Eigen-Inference for Energy Estimation of Multiple Sources

Romain Couillet, *Student Member, IEEE,* Jack W. Silverstein, Zhidong Bai, and Mérouane Debbah, *Senior Member, IEEE*

***Abstract*—In this paper, a new method is introduced to blindly estimate the transmit power of multiple signal sources in multi-antenna fading channels, when the number of sensing devices and the number of available samples are sufficiently large compared to the number of sources. Recent advances in the field of large dimensional random matrix theory are used that result in a simple and computationally efficient consistent estimator of the power of each source. A criterion to determine the minimum number of sensors and the minimum number of samples required to achieve source separation is then introduced. Simulations are performed that corroborate the theoretical claims and show that the proposed power estimator largely outperforms alternative power inference techniques.**

***Index Terms*—Cognitive radio, G-estimation, power estimation, random matrix theory, statistical inference.**

I. Introduction

At a time when radio resources become scarce, the alternative offered by cognitive radios [1] is gaining more and more interest. A *cognitive* (or *flexible*) wireless network is a set of opportunistic entities, referred to as the *secondary* network, that benefit from unused spectrum resources to establish communication while generating little or even no interference to the licensed networks, collectively referred to as the *primary* network. This is achieved by letting the secondary devices sense the communication channel for the presence of active transmissions and exchange the collected information among the secondary network, in order to perform optimal decisions on the opportunistic communication strategy to apply. The difficulty for the secondary network does not lie in the detection of downlink transmissions from fixed access points to licensed mobile users in the primary network, but rather in the reliable detection of the uplink transmissions from the mobile licensed users to the primary access points. If, in addition to detecting active transmissions, the secondary devices can, at all time, detect the exact number of primary mobile sources and evaluate the power used by every individual source, the transmission policy in the secondary network can be accurately and dynamically adapted. An example of use is found in the recent development of femtocells, i.e., small area cells that operate indoors by overlaying the spectrum licensed to outdoors macrocells. Closed access femtocells have the capabilities to self-organize and to dynamically access spectrum resources [2]-[3]; specifically, the first requirement of a femtocell is to minimally interfere the overlaid licensed macrocell network, while simultaneously trying to optimize transmission data rates within the femtocell. This requires that the femtocells be constantly aware of the outdoor activity of the macrocell mobile users. As such, macrocell-femtocell networks are cognitive wireless networks in which the established macrocell network is seen as the primary network, while the femtocell network plays the role of the opportunistic secondary network. In [4], the achievable rates of a two-tier macrocell-femtocell network are derived in the very general case where all entities in the networks are embedded with multiple antennas. The optimal coverage of the secondary networks is computed under several assumptions on the side information available at the femtocells. Among these assumptions, [4] supposes that the femtocells have perfect knowledge of the distances to the macrocell user equipments. This last assumption suggests that the femtocells have a means, either global positioning system or some sort of detection mechanism, to perfectly evaluate the distances to the active primary users. In the present work, we address the problem of the estimation of the distance of the secondary network to each primary user or, more exactly, the problem of the estimation of the individual source transmit powers. We provide a framework for the secondary network (i) to identify the number of primary sources, (ii) to determine the number of transmit antennas for every source and (iii) to estimate the transmit power from each source.

The difficulty of estimating transmit powers lies in the little information known *a priori* by the secondary network: the transmitted data and the transmission channels are usually inaccessible. This has motivated much work in the direction of blind signal source detection methods, based on the Neyman-Pearson test in Gaussian channels [5], Rayleigh fading channels [6], multiple antenna channels [7] and large dimensional multi-antenna channels [8], but these successive works are designed to answer a binary hypothesis test on the presence or absence of a signal source. Alternatively, in [9], a method is derived to separate signal sources and estimate the number of those sources. To solve the harder problem of power inference, it is necessary to assume that the amount of sensors in the secondary network is larger than the number of active

R. Couillet and M. Debbah are with the Alcatel-Lucent Chair on Flexible Radio, SUPÉLEC, Gif sur Yvette, 91192, Plateau de Moulon, 3, Rue Joliot-Curie, France e-mail: {romain.couillet, merouane.debbah}@supelec.fr. M. Debbah's work is supported by the European Commission, FP7 Network of Excellence in Wireless Communications NEWCOM++ and the French ANR Project SESAME.

J. W. Silverstein is with the Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205, jack@math.ncsu.edu. J. W. Silverstein's work is supported by the U.S. Army Research Office, Grant W911NF-09-1-0266.

Z. Bai is with the KLAS MOE & School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China, baizd@nenu.edu.cn. Z. Bai's work is supported by the NSF China, Grant 10871036, the NUS, Grant R-155-000-079-112 and R-155-000-096-646.

sources, e.g., individual secondary users are equipped with many antennas, or a large number of secondary users, each of them equipped with few antennas, collect their received data via a central backbone; this assumption is valid in the context of femtocells that can communicate through wired private or public networks. The condition on the number of sensors allows one to model the multi-dimensional channel $\mathbf{H}$ from the joint primary sources to the secondary users, the joint source transmit data $\mathbf{X}$ and the additive received noise $\mathbf{W}$ as large dimensional random matrices with independent entries (no specific matrix size definition is required at this point). Denoting $\mathbf{P}$ a diagonal matrix whose entries are the source powers with multiplicities the number of transmit antennas of each user, the power detection problem boils down to estimating the entries of $\mathbf{P}$ from the sole knowledge of the received data matrix $\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \mathbf{W}$, as all system dimensions (number of antennas per transmit source, number of sensors, number of available samples) are large. If the available samples largely outnumber the sensors (of several orders of magnitude), and the number of sensors are much larger than the number of transmit antennas, the strong law of large numbers ensures that the diagonal entries of $\mathbf{P}$ can be retrieved directly from the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, and the problem is immediately solved. When all dimensions are large but are of the same order of magnitude, the law of large numbers no longer applies and one has to consider results from the theory of large dimensional random matrices, e.g., [10], used in the present article to derive the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ as a function of $\mathbf{P}$. To this day and to the best of our knowledge, no computationally-efficient *consistent* estimator for the entries of $\mathbf{P}$ has been proposed.[1] Among the existing techniques are discretization and convex optimization strategies [11], [12], which tend to directly invert the result from [10] (although an explicit inverse was not available at that time), and moment-based approaches [13], [14], which use the empirical moments of the eigenvalue distribution $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ to infer the entries of $\mathbf{P}$. Some of these moment-based methods are computationally cheap, but provide in general consistent estimators of the moments of the eigenvalue distribution of $\mathbf{P}$, instead of estimators of the sought powers. These techniques are therefore expected to perform worse than methods that would fully exploit the eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, and not only a few moments of the distribution. This problem is successfully addressed in [15] for the simpler *sample covariance matrix* model $\mathbf{Y}' = \mathbf{P}^{\frac{1}{2}}\mathbf{X}$, where strongly consistent estimators for the individual entries of $\mathbf{P}$ are provided, which are based on the full eigenvalue distribution of $\mathbf{X}^{\mathsf{H}}\mathbf{P}\mathbf{X}$.

The present work generalizes this result to infer the entries of $\mathbf{P}$ from the observed matrix $\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \mathbf{W}$. The novel estimator proposed here is strongly consistent with respect to growing number of sensors, sources and samples, has a very compact form, is computationally efficient and is shown in simulations to largely outperform alternative approaches, such as moment-based methods. The estimator is moreover robust to small system dimensions. We specifically show that, if the number of sensing entities is larger than the number of active transmitters in the primary network, it is possible to evaluate both the exact number of transmitters and their respective transmit powers (and, for that matter, the number of transmit antennas per source can also be estimated). Otherwise, ambiguous scenarios might arise where multiple transmitters may be confused as a single transmitter with estimated transmit power the average of the true transmit powers of these transmitters. Additionally, we provide an expression of the minimum number of sensors required to separate transmit sources of similar power.

[1] an estimator $\hat{P}_i$ of the i^{th} entry P_i of $\mathbf{P}$ is said to be *consistent* if $\hat{P}_i - P_i \to 0$ almost surely when the relevant system dimensions grow large.

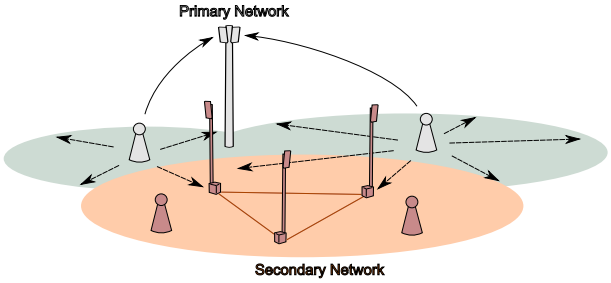

Fig. 1. A cognitive radio network

The remainder of this paper is structured as follows: Section II introduces the system model. In Section III, we study the asymptotic spectrum of the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. In Section IV, the novel power estimator is derived. Section V provides simulation results. Section VI concludes this work.

Notations: In the following, boldface lower case symbols represent vectors, capital boldface characters denote matrices ($\mathbf{I}_N$ is the size-N identity matrix). The transpose and Hermitian transpose operators are denoted $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$, respectively. We denote by $\mathbb{C}^+$ the set $\{z \in \mathbb{C}, \Im[z] > 0\}$ and by $\mathbb{C}^-$ the set $\{z \in \mathbb{C}, \Im[z] < 0\}$. The left-limit in x of a function f is denoted $f(x-)$.

II. System Model

Consider a wireless (primary) network in which K entities are transmitting data simultaneously on the same frequency resource. Transmitter $k \in \{1, \ldots, K\}$ has transmission power P_k and is equipped with n_k antennas. We denote $n = \sum_{k=1}^{K} n_k$ the total number of transmit antennas of the primary network. Consider also a secondary network composed of a total of N, $N \geq n$, sensing devices (they may be N single antenna devices or multiple devices embedded with multiple antennas whose sum equals N); we shall refer to the N sensors collectively as *the receiver*. This scenario is depicted in Figure 1. To ensure that every sensor in the secondary network, e.g., in a femtocell, roughly captures the same amount of energy from a given transmitter, we need to assume that the respective transmitter-sensor distances are alike. This is a realistic assumption for a in-house femtocell network. Denote $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ the multiple antenna channel matrix between transmitter k and the receiver. We assume that the entries of $\sqrt{N}\mathbf{H}_k$ are independent and identically distributed with zero mean, unit variance and finite fourth order moment. At time instant m, transmitter k emits the multi-antenna signal $\mathbf{x}_k^{(m)} \in \mathbb{C}^{n_k}$, with entries assumed to be independent and

identically distributed of zero mean, unit variance and finite fourth order moment. Assume further that at time instant m the receive signal is impaired by additive white noise with entries of zero mean, variance σ^2 and finite fourth order moment on every sensor; we denote $\sigma\mathbf{w}^{(m)} \in \mathbb{C}^N$ the receive noise vector where the entries of $\mathbf{w}_k^{(m)}$ have unit variance. At time m, the receiver therefore senses the signal $\mathbf{y}^{(m)} \in \mathbb{C}^N$ defined as

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}. \quad (1)$$

Assuming the channel fading coefficients are constant over at least M consecutive sampling periods, by concatenating M successive signal realizations into $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$, we have

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} \quad (2)$$

where $\mathbf{X}_k = [\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k\times M}$, for every k, and $\mathbf{W} = [\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(M)}] \in \mathbb{C}^{N\times M}$. This can be further rewritten as

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W} \quad (3)$$

where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is diagonal with first n_1 entries P_1, subsequent n_2 entries P_2, ..., last n_K entries P_K, $\mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_K] \in \mathbb{C}^{N\times n}$ and $\mathbf{X} = [\mathbf{X}_1^{\mathsf{T}}, \ldots, \mathbf{X}_K^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{n\times M}$. By convention, we shall assume $P_1 \leq \ldots \leq P_K$.

Remark 1: The statement that $\sqrt{N}\mathbf{H}$, $\mathbf{X}$ and $\mathbf{W}$ have independent entries of finite fourth order moment is meant to provide as loose assumptions as possible on the channel, signal and noise properties. In the simulations of Section V, the entries of $\mathbf{H}$, $\mathbf{W}$ are taken Gaussian. Nonetheless, according to our assumptions, the entries of $\mathbf{X}$ need not be identically distributed, but may originate from a maximum of K distinct distributions. This translates the realistic assumption that different data sources may use different symbol constellations (e.g., M-QAM, M-PSK); the finite fourth moment assumption is obviously verified for finite constellations.

Our objective is to infer the values of the powers $P_1, \ldots, P_K$ from the realization of the random matrix $\mathbf{Y}$. This is the subject of Section IV. In the sequel, we introduce tools from large dimensional random matrix theory and we provide a thorough analysis of the eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ as N, n and M grow large at the same rate.

III. Spectral analysis

We start by analyzing the eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when n, N and M grow large at a similar rate. This is a fundamental prior step to the proper estimation of $P_1, \ldots, P_K$.

A. Limiting spectrum of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

We first define the Stieltjes transform of a (cumulative) distribution function.

Definition 1: Let F be a distribution function. For $z \in \mathbb{C} \setminus \mathbb{R}^+$, the *Stieltjes transform* $m(z)$ of F is defined as

$$m(z) = \int \frac{1}{t-z} dF(t). \quad (4)$$

For $x \in \mathbb{R}$ a continuity point of F, we have the *inverse Stieltjes transform* formula

$$F(x) = \frac{1}{\pi} \lim_{y\to 0^+} \int_{-\infty}^{x} \Im[m(t+iy)]dt. \quad (5)$$

In this section, we prove the following result

Theorem 1: Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, with $\mathbf{Y}$ defined as in (3). Then, for M, N, n growing large with limit ratios $M/N \to c$, $N/n_k \to c_k$, $0 < c, c_1, \ldots, c_K < \infty$, the eigenvalue distribution function $F^{\mathbf{B}_N}$ of $\mathbf{B}_N$, referred to as the *empirical spectral distribution* (e.s.d.) of $\mathbf{B}_N$, converges almost surely to the deterministic distribution function F, referred to as the *limit spectral distribution* (l.s.d.) of $\mathbf{B}_N$, whose Stieltjes transform $m_F(z)$ satisfies, for $z \in \mathbb{C}^+$,

$$m_F(z) = cm_{\underline{F}}(z) + (c-1)\frac{1}{z} \quad (6)$$

where $m_{\underline{F}}(z)$ is the unique solution with positive imaginary part of the implicit equation in $m_{\underline{F}}$

$$\frac{1}{m_{\underline{F}}} = -\sigma^2 + \frac{1}{f} - \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1+P_k f} \quad (7)$$

in which we denoted f the value

$$f = (1-c)m_{\underline{F}} - czm_{\underline{F}}{}^2. \quad (8)$$

The rest of this section is dedicated to the proof of Theorem 1. First remark that (3) can be further simplified into

$$\mathbf{Y} = \begin{pmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix}. \quad (9)$$

Appending $\mathbf{Y} \in \mathbb{C}^{N\times M}$ into the larger matrix $\underline{\mathbf{Y}} \in \mathbb{C}^{(N+n)\times M}$

$$\underline{\mathbf{Y}} = \begin{pmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix}, \quad (10)$$

we recognize that $\frac{1}{M}\underline{\mathbf{Y}}\underline{\mathbf{Y}}^{\mathsf{H}}$ is a *sample covariance matrix*, for which the *population covariance matrix* $\begin{pmatrix}\mathbf{H}\mathbf{P}\mathbf{H}^{\mathsf{H}}+\sigma^2\mathbf{I}_N & 0\\ 0 & 0\end{pmatrix}$ is non-deterministic and the random matrix $\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix}$ has independent (non-necessarily identically distributed) entries with zero mean and variance 1.

At this point, we need the following result,

Proposition 1: Let $\mathbf{Z}_n \in \mathbb{C}^{N\times n}$ have complex independent entries of zero mean, unit variance and finite $2+\varepsilon$ order moment, for some $\varepsilon > 0$, and $\mathbf{T}_n \in \mathbb{R}^{n\times n}$ be Hermitian with e.s.d. converging almost surely to T, as $N \to \infty$. Let $\mathbf{A}_n = \frac{1}{N}\mathbf{Z}_n\mathbf{T}_n\mathbf{Z}_n^{\mathsf{H}}$. Then, as $n, N \to \infty$, $N/n \to c > 0$, the eigenvalue distribution of $\mathbf{A}_n$ converges weakly and almost surely to the distribution function A with Stieltjes transform $m_A(z)$, $z \in \mathbb{C}^+$, being the unique solution with positive imaginary part of the equation in m_A

$$z = -\frac{1}{m_A} + \frac{1}{c}\int \frac{t}{1+tm_A} dT(t). \quad (11)$$

Proof: The proof originates from Theorem 4.1 of [16] that states that, under the hypotheses of Proposition 1, the eigenvalue distribution of $\mathbf{A}_n$ converges weakly to some distribution function A whose Stieltjes transform $m_A(z)$ is a function of the Stieltjes transform of $m_T(z)$ and c only; $m_A(z)$

is explicitly given by (4.4.4) of [16]. Now, in the special case where $\mathbf{Z}_n$ has independent and identically distributed (i.i.d.) entries of zero mean, unit variance and finite $2+\varepsilon$ moment, [10] and Theorem 4.3 of [16] show that $m_A(z)$ satisfies (11). But then, since $m_A(z)$ is only a function of c and T regardless of the distribution of the independent entries of $\mathbf{Z}_n$, $m_A(z)$ that solves (11) is the Stieltjes transform of A for the more general case. ∎

Note that Proposition 1 can be equally stated when $z \in \mathbb{C}^-$. In that case, $m_A(z)$ is the unique solution of (11) with negative imaginary part.

From Proposition 1, since $\mathbf{H}$ has independent entries with finite fourth order moment, we have that the e.s.d. of $\mathbf{HPH}^{\mathsf{H}}$ converges weakly and almost surely to a limit distribution G as $N, n_1, \ldots, n_K \to \infty$ with, $N/n_k \to c_k > 0$. For $z \in \mathbb{C}^+$, the Stieltjes transform $m_G(z)$ of G is the unique solution with positive imaginary part of the equation in m_G,

$$z = -\frac{1}{m_G} + \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1 + P_k m_G}. \tag{12}$$

The almost sure convergence of the e.s.d. of $\mathbf{HPH}^{\mathsf{H}}$ ensures the almost sure convergence of the e.s.d. of the matrix $\begin{pmatrix} \mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$. Since $m_G(z)$ evaluated at $z \in \mathbb{C}^+$ is the Stieltjes transform of the l.s.d. of $\mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$ evaluated at $z + \sigma^2$, adding n zero eigenvalues, we finally have that the e.s.d. of $\begin{pmatrix} \mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$ tends almost surely to a distribution H whose Stieltjes transform $m_H(z)$ satisfies

$$m_H(z) = \frac{c_0}{1+c_0} m_G(z - \sigma^2) - \frac{1}{1+c_0} \frac{1}{z}, \tag{13}$$

for $z \in \mathbb{C}^+$, where we denoted c_0 the limit of the ratio N/n, i.e., $c_0 = (c_1^{-1} + \ldots + c_K^{-1})^{-1}$.

As a consequence, the sample covariance matrix $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ has a population covariance matrix which is not deterministic but whose e.s.d. has an almost sure limit for increasing dimensions. Since $\mathbf{X}$ and $\mathbf{W}$ have entries with finite fourth order moment, we can again apply Proposition 1, and we have that the e.s.d. of $\underline{\mathbf{B}}_N \triangleq \frac{1}{M}\mathbf{Y}^{\mathsf{H}}\mathbf{Y}$ converges almost surely to the limit $\underline{F}$ whose Stieltjes transform $m_{\underline{F}}(z)$ is the unique solution in $\mathbb{C}^+$ of the equation in $m_{\underline{F}}$

$$z = -\frac{1}{m_{\underline{F}}} + \frac{1}{c}\left(1 + \frac{1}{c_0}\right) \int \frac{t}{1 + t m_{\underline{F}}} dH(t) \tag{14}$$

$$= -\frac{1}{m_{\underline{F}}} + \frac{1 + \frac{1}{c_0}}{c m_{\underline{F}}} \left[1 - \frac{1}{m_{\underline{F}}} m_H\left(-\frac{1}{m_{\underline{F}}}\right)\right] \tag{15}$$

for all $z \in \mathbb{C}^+$.

For $z \in \mathbb{C}^+$, $m_{\underline{F}}(z) \in \mathbb{C}^+$. Therefore $-1/m_{\underline{F}}(z) \in \mathbb{C}^+$ and one can evaluate (13) at $-1/m_{\underline{F}}(z)$. Combining (13) and (15), we then have

$$z = -\frac{1}{c} \frac{1}{m_{\underline{F}}(z)^2} m_G\left(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2\right) + \left(\frac{1}{c} - 1\right) \frac{1}{m_{\underline{F}}(z)}, \tag{16}$$

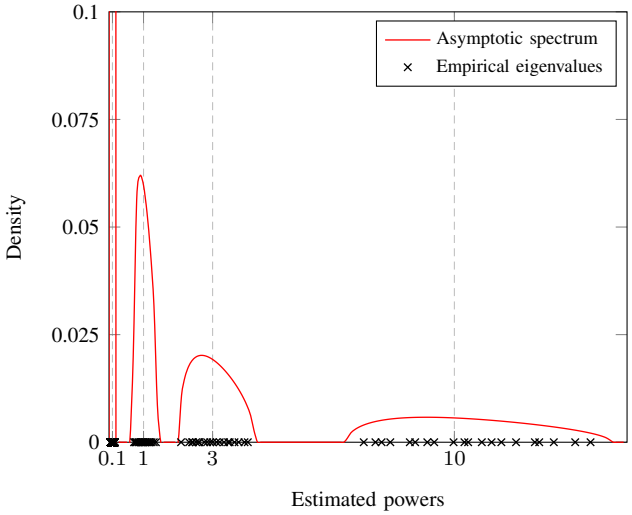

Fig. 2. Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $c_0 = 10$, $c = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

where, according to (12), $m_G(-1/m_{\underline{F}}(z) - \sigma^2)$ satisfies

$$\frac{1}{m_{\underline{F}}(z)} = -\sigma^2 + \frac{1}{m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2)} - \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1 + P_k m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2)}. \tag{17}$$

Together with (16), this is exactly (7), with $f(z) = m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2) = (1-c)m_{\underline{F}}(z) - czm_{\underline{F}}(z)^2$.

Since the eigenvalues of the matrices $\mathbf{B}_N$ and $\underline{\mathbf{B}}_N$ only differ by $M-N$ zeros, we also have that the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$ satisfies

$$m_F(z) = c m_{\underline{F}}(z) + (c-1)\frac{1}{z}. \tag{18}$$

This completes the proof of Theorem 1. For further usage, notice here that (18) provides a simplified expression for $m_G(-1/m_{\underline{F}}(z) - \sigma^2)$. Indeed we have,

$$m_G(-1/m_{\underline{F}}(z) - \sigma^2) = -z m_F(z) m_{\underline{F}}(z). \tag{19}$$

Therefore, the support of the (almost sure) l.s.d. F of $\mathbf{B}_N$ can be evaluated as follows: for any $z \in \mathbb{C}^+$, $m_F(z)$ is given by (6), in which $m_{\underline{F}}(z)$ is solution of (7); the inverse Stieltjes transform formula (5) allows then to evaluate F from $m_F(z)$, for values of z spanning over the set $\{z = x + iy, x > 0\}$ and y small. This is depicted in Figure 2, when $\mathbf{P}$ has three distinct values $P_1 = 1$, $P_2 = 3$, $P_3 = 10$ and $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$, as well as in Figure 3 for the same setup but with $P_3 = 5$.

Two remarks on Figures 2 and 3 are of fundamental importance to the following. First, it appears that the asymptotic l.s.d. F of $\mathbf{B}_N$ is compactly supported and divided into up to $K+1$ disjoint compact intervals, which we further refer to as *clusters*. Each cluster can be mapped onto one or many values in the set $\{\sigma^2, P_1, \ldots, P_K\}$. For instance, in Figure 3, the first cluster is mapped to σ^2, the second cluster to P_1 and the third cluster to the set $\{P_2, P_3\}$. Depending on the ratios c

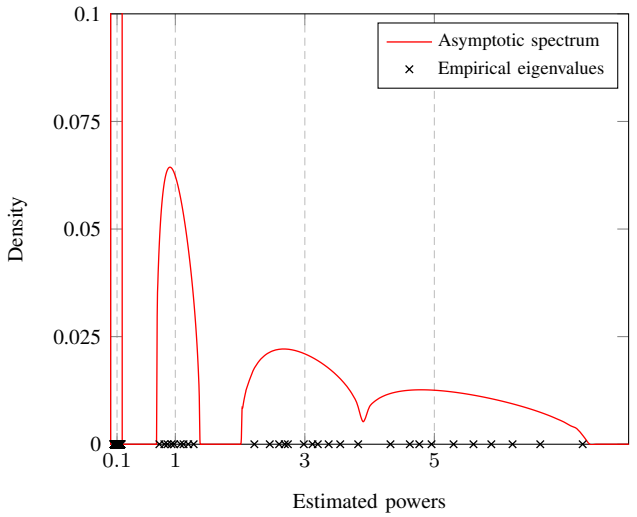

Fig. 3. Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 5$, $n_1 = n_2 = n_3$, $c_0 = 10$, $c = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

and c_0 and on the particular values taken by $P_1, \ldots, P_K$ and σ^2, these clusters are either thin disjoint compact intervals, as in Figure 2, or they may overlap to generate larger compact intervals, as in Figure 3. We shall see, as is in fact required by the law of large numbers, that for increasing c and c_0, the asymptotic spectrum tends to be divided into thinner and thinner clusters. The inference technique proposed hereafter relies on the separability of the clusters associated to each P_i and to σ^2. Precisely, to be able to derive a consistent estimate of the transmitted power P_k, the cluster associated to P_k in F, number it cluster k_F, must be distinct from the neighboring clusters $(k-1)_F$ and $(k+1)_F$, associated to P_{k-1} and P_{k+1} respectively (when they exist), and also distinct from cluster 1 in F associated to σ^2. As such, in the scenario of Figure 3, our method will be able to provide a consistent estimate for P_1, but (so far) will not succeed in providing a consistent estimate for either P_2 or P_3, since $2_F = 3_F$. We shall see that a consistent estimate for $(P_2 + P_3)/2$ is accessible though. Secondly, notice that the empirical eigenvalues of $\mathbf{B}_N$ are all inside the asymptotic clusters and, most importantly, in the case where cluster k_F is distinct from 1, $(k-1)_F$ and $(k+1)_F$, observe that the number of eigenvalues in cluster k_F is exactly n_k. This fact is referred to as *exact separation*. The exact separation for the current model originates from a direct application of the exact separation for the sample covariance matrix proven in [17] and is provided here in Theorem 3. This is further discussed in the subsequent sections.

B. Condition for separability

In the following, we are interested in estimating consistently the power P_k for a given fixed $k \in \{1, \ldots, K\}$. We recall that consistency means here that, as all system dimensions grow large with finite asymptotic ratios, the difference $\hat{P}_k - P_k$ between the estimate $\hat{P}_k$ of P_k and P_k itself converges to zero with probability one. As previously mentioned, we show by construction in Section IV that such an estimate is only achievable if the cluster mapped to P_k in F is disjoint from all other clusters. The purpose of the present section is to provide sufficient conditions for cluster separability.

To ensure that cluster k_F (associated to P_k in F) is distinct from cluster 1 (associated to σ^2) and clusters i_F, $i \neq k$, (associated to all other P_i), we assume now and for the rest of this article that the following conditions are fulfilled: (i) k satisfies Assumption 1, given as follows

Assumption 1:

$$\sum_{r=1}^{K} \frac{1}{c_r} \frac{(P_r m_{G,k})^2}{(1 + P_r m_{G,k})^2} < 1 \tag{20}$$

$$\sum_{r=1}^{K} \frac{1}{c_r} \frac{(P_r m_{G,k+1})^2}{(1 + P_r m_{G,k+1})^2} < 1 \tag{21}$$

with $m_{G,1}, \ldots, m_{G,K}$ the K real solutions to the equation in m_G,

$$\sum_{r=1}^{K} \frac{1}{c_r} \frac{(P_r m_G)^3}{(1 + P_r m_G)^3} = 1 \tag{22}$$

with the convention $m_{G,K+1} = 0$,
and (ii) k satisfies Assumption 2 as follows,

Assumption 2: Denoting, for $j \in \{1, \ldots, K\}$,

$$j_G \triangleq \#\{i \leq j \mid i \text{ satisfies Assumption 1}\}, \tag{23}$$

$$\begin{aligned}
&\frac{1-c_0}{c_0} \frac{(\sigma^2 m_{\underline{F},k_G})^2}{(1+\sigma^2 m_{\underline{F},k_G})^2} \\
&+ \sum_{r=1}^{k_G-1} \frac{1}{c_r} \frac{(x_{G,r}^+ + \sigma^2)^2 m_{\underline{F},k_G}^2}{(1+(x_{G,r}^+ + \sigma^2) m_{\underline{F},k_G})^2} \\
&+ \sum_{r=k_G}^{K_G} \frac{1}{c_r} \frac{(x_{G,r}^- + \sigma^2)^2 m_{\underline{F},k_G}^2}{(1+(x_{G,r}^- + \sigma^2) m_{\underline{F},k_G})^2} < c
\end{aligned} \tag{24}$$

$$\begin{aligned}
&\frac{1-c_0}{c_0} \frac{(\sigma^2 m_{\underline{F},k_G+1})^2}{(1+\sigma^2 m_{\underline{F},k_G+1})^2} \\
&+ \sum_{r=1}^{k_G} \frac{1}{c_r} \frac{(x_{G,r}^+ + \sigma^2)^2 m_{\underline{F},k_G+1}^2}{(1+(x_{G,r}^+ + \sigma^2) m_{\underline{F},k_G+1})^2} \\
&+ \sum_{r=k_G+1}^{K_G} \frac{1}{c_r} \frac{(x_{G,r}^- + \sigma^2)^2 m_{\underline{F},k_G+1}^2}{(1+(x_{G,r}^- + \sigma^2) m_{\underline{F},k_G+1})^2} < c
\end{aligned} \tag{25}$$

where $x_{G,i}^-, x_{G,i}^+$, $i \in \{1, \ldots, K_G\}$, are defined by

$$x_{G,i}^- = -\frac{1}{m_{G,i}^-} + \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r}{1 + P_r m_{G,i}^-} \tag{26}$$

$$x_{G,i}^+ = -\frac{1}{m_{G,i}^+} + \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r}{1 + P_r m_{G,i}^+} \tag{27}$$

with $m_{G,1}^-, m_{G,1}^+, \ldots, m_{G,K_G}^-, m_{G,K_G}^+$ the $2K_G$ real roots of (20), and $m_{\underline{F},j}$, $j \in \{1, \ldots, K_G + 1\}$, the j-th real root (in

increasing order) of the equation in $m_{\underline{F}}$

$$\frac{1-c_0}{c_0}\frac{(\sigma^2 m_{\underline{F}})^3}{(1+\sigma^2 m_{\underline{F}})^3}+\sum_{r=1}^{j-1}\frac{1}{c_r}\frac{(x_{G,r}^{+}+\sigma^2)^3 {m_{\underline{F}}}^3}{(1+(x_{G,r}^{+}+\sigma^2)m_{\underline{F}})^3} + \sum_{r=j}^{K_G}\frac{1}{c_r}\frac{(x_{G,r}^{-}+\sigma^2)^3 {m_{\underline{F}}}^3}{(1+(x_{G,r}^{-}+\sigma^2)m_{\underline{F}})^3}=c. \tag{28}$$

Although difficult to fathom at this point of the article, the above assumptions will be clarified in the subsequent sections. We give hereafter a short intuitive explanation of the role of every condition.

Assumption 1 is a necessary and sufficient condition for cluster k_G, that we define as the cluster associated to P_k in G (the l.s.d. of $\mathbf{HPH}^{\mathsf{H}}$), to be distinct from the clusters $(k-1)_G$ and $(k+1)_G$, associated to P_{k-1} and P_{k+1} in G, respectively. Note that we implicitly assume a unique mapping between the P_i and clusters in G; this statement will be made more rigorous in subsequent sections. Assumption 1 only deals with the inner $\mathbf{HPH}^{\mathsf{H}}$ covariance matrix properties and ensures specifically that the powers to be estimated differ sufficiently from one another for our method to be able to resolve them. Note that, if $P_1,\ldots,P_K$ are scaled by a common constant, then the solutions of (22) are scaled by the inverse of this constant; the separability condition is then a function of $P_2/P_1,\ldots,P_K/P_1$ and of the ratios $c_1,\ldots,c_K$ only. In Figure 4, we depict the critical ratio c_0 above which Assumption 1 is satisfied for all k, when $K=2$ and $c_1=c_2$, as a function of P_1/P_2, i.e., the critical ratio c_0 above which the two clusters associated to P_1 and P_2 in G are disjoint. Observe that, as P_1 gets close to P_2, c_0 increases fast; therefore, to be able to separate power values with ratio close to one, an extremely large number of sensors is required. In Figure 5, the case $K=3$ is considered with $c_1=c_2=c_3$, $c_0=10$, and we let P_2/P_1 and P_3/P_1 vary; this situation corresponds to the scenarios previously depicted in Figures 2 and 3. Note that the triplet $(P_1,P_2,P_3)=(1,3,5)$ is slightly outside the region that satisfies Assumption 1, and then, for this c_0, not all the clusters of G (and therefore of F) are disjoint, as confirmed by Figure 3. As for the triplet $(1,3,10)$, it clearly lies inside the region that satisfies Assumption 1, which is sufficient to ensure the separability of the clusters in G, but not enough though to ensure the separability of the clusters in F.

Assumption 2 deals with the complete $\mathbf{B}_N$ matrix model. It is however a non-necessary but sufficient condition so that cluster k_F, associated to P_k in F, be distinct from clusters $(k-1)_F$, $(k+1)_F$ and 1 (cluster 1 being associated to σ^2). The exact necessary and sufficient condition will be stated further in the next sections; however, the latter is not exploitable as is, and Assumption 2 will be shown to be an appropriate substitute. Assumption 2 is concerned with the value of c necessary to avoid (i) cluster k_G (associated to P_k in G) to further spread on the clusters k_G-1 and k_G+1 associated to P_{k-1} and P_{k+1} and, more importantly, to avoid (ii) cluster 1 associated to σ^2 in F to merge with cluster k_F. As shall become evident in the next sections, when σ^2 is large, the tendency is for the cluster associated to σ^2 to become large and spread over the clusters associated to P_1, then P_2 etc. To counter this effect, one must increase c, i.e., take more signal samples. Figure 6 depicts the critical ratio c that satisfies Assumption 2 as a function of σ^2, in the case $K=3$, $(P_1,P_2,P_3)=(1,3,10)$, $c_0=10$, $c_1=c_2=c_3$. Notice that, in the case $c=10$, below $\sigma^2\simeq 1$, it is possible to separate all clusters, which is compliant with Figure 2 where $\sigma^2=0.1$.

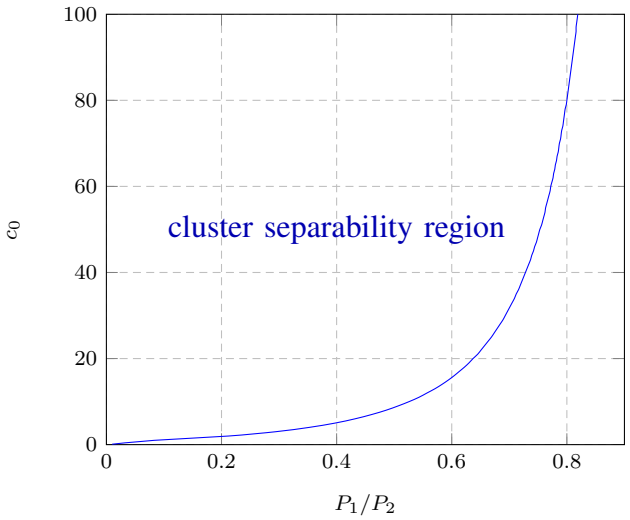

Fig. 4. Limiting ratio c_0 to ensure separability of (P_1,P_2), $P_1\le P_2$, $K=2$, $c_1=c_2$.

As a consequence, under the assumption (proved later) that our proposed method cannot perform consistent power estimation when the cluster separability conditions are not met, we have two first conclusions:

- if one desires to increase the sensitivity of the estimator, i.e., to be able to separate two sources of close transmit powers, one needs to increase the number of sensors (by increasing c_0),
- if one desires to detect and reliably estimate power sources in a noisy environment, one needs to increase the number of sensed samples (by increasing c).

In the subsequent section, we study the properties of the asymptotic spectrum of $\mathbf{HPH}^{\mathsf{H}}$ and $\underline{\mathbf{B}}_N$ in more details. These properties will lead to an explanation for Assumptions 1 and 2. Under those assumptions, we shall then derive our novel power estimator.

IV. Multi-source power inference

In this section, we prove our main result,

Theorem 2: Let $\mathbf{B}_N\in\mathbb{C}^{N\times N}$ be defined as in Theorem 1, and $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_N)$, $\lambda_1\le\ldots\le\lambda_N$, be the vector of the ordered eigenvalues of $\mathbf{B}_N$. Further assume that the limiting ratios c_0, $c_1,\ldots,c_K$, c and $\mathbf{P}$ are such that Assumptions 1 and 2 are fulfilled for some $k\in\{1,\ldots,K\}$. Then, as N, n, M grow large, we have

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0 \tag{29}$$

where the estimate $\hat{P}_k$ is given by

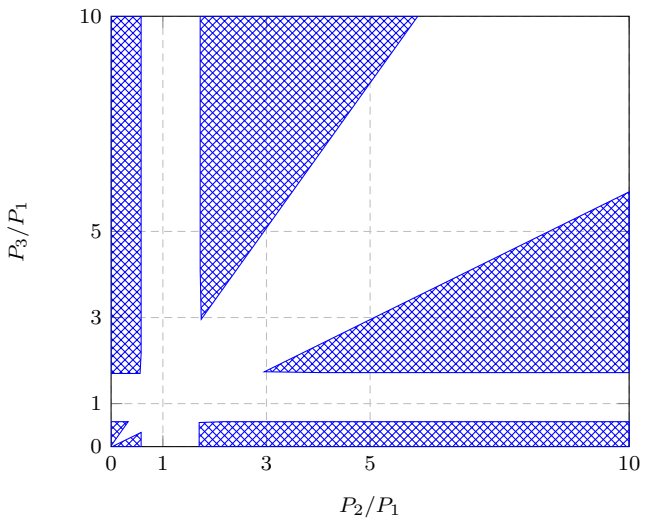

Fig. 5. Subset of (P_1, P_2, P_3) that fulfills Assumption 1 $K = 3$, $c_1 = c_2 = c_3$, for $c_0 = 10$, in crosshatched pattern.

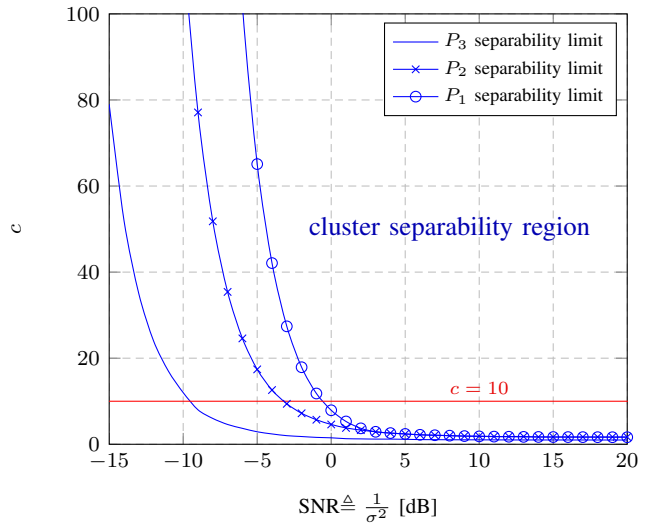

Fig. 6. Limiting ratio c as a function of σ^2 to ensure consistent estimation of $P_1 = 1$, $P_2 = 3$ and $P_3 = 10$, $c_0 = 10$, $c_1 = c_2 = c_3$.

- if $M \neq N$,

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i \in \mathcal{N}_k} (\eta_i - \mu_i), \tag{30}$$

- if $M = N$,

$$\hat{P}_k = \frac{N}{n_k(N-n)} \sum_{i \in \mathcal{N}_k} \left(\sum_{j=1}^{N} \frac{\eta_i}{(\lambda_j - \eta_i)^2} \right)^{-1}, \tag{31}$$

in which $\mathcal{N}_k = \{\sum_{i=1}^{k-1} n_i + 1, \ldots, \sum_{i=1}^{k} n_i\}$, $(\eta_1, \ldots, \eta_N)$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $(\mu_1, \ldots, \mu_N)$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Remark 2: We immediately notice that, if $N < n$, the powers $P_1, \ldots, P_l$, with l the largest integer such that $N - \sum_{i=l}^{K} n_i < 0$, cannot be estimated.

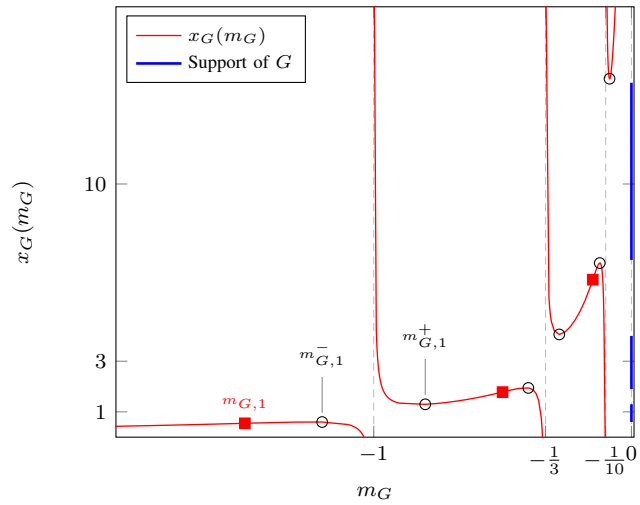

Fig. 7. $x_G(m_G)$ for m_G real, $\mathbf{P}$ diagonal composed of three evenly weighted masses in 1, 3 and 10. Local extrema are marked in circles, inflexion points are marked in squares.

The approach pursued to prove Theorem 2 relies strongly on the original idea of [15]. From Cauchy's integration formula [18],

$$\begin{aligned} P_k &= \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{\omega}{P_k - \omega} d\omega \\ &= c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{P_r - \omega} d\omega \end{aligned} \tag{32}$$

for any negatively oriented contour $\mathcal{C}_k \subset \mathbb{C}$, such that P_k is contained in the surface described by the contour, while for every $i \neq k$, P_i is outside this surface. The strategy is then the following: we first propose a convenient integration contour $\mathcal{C}_k$ which is parametrized by a function of the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$. We proceed to a variable change in (32) to express P_k as a function of $m_F(z)$. We then evaluate the complex integral resulting from replacing the limiting $m_F(z)$ in (32) by its empirical counterpart $\hat{m}_F(z) = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. This new integral, whose value we name $\hat{P}_k$, is shown to be almost surely equal to P_k in the large N limit. It then suffices to evaluate $\hat{P}_k$, which is just a matter of residue calculus [18].

We start by determining the integration contour $\mathcal{C}_k$. For this, we first need to study the distributions G and F in more details.

A. Properties of G and F

Let us introduce the following result on the l.s.d. of sample covariance matrices, borrowed from [19]

Proposition 2: Let $\mathbf{A}_n$ be defined as in Proposition 1. Then the almost sure limiting Stieltjes transform $m_A(z)$ of the e.s.d. of $\mathbf{A}_n$, $z \in \mathbb{C}^+$ admits a limit $m_A^\circ(x)$ when $z \to x \in \mathbb{R}^*$. If x is inside the support of A, then $m_A^\circ(x)$ is the only solution with positive imaginary part of the equation $x_A(m) = x$ in the variable m, with $x_A(m)$ defined, for $-1/m$ outside the

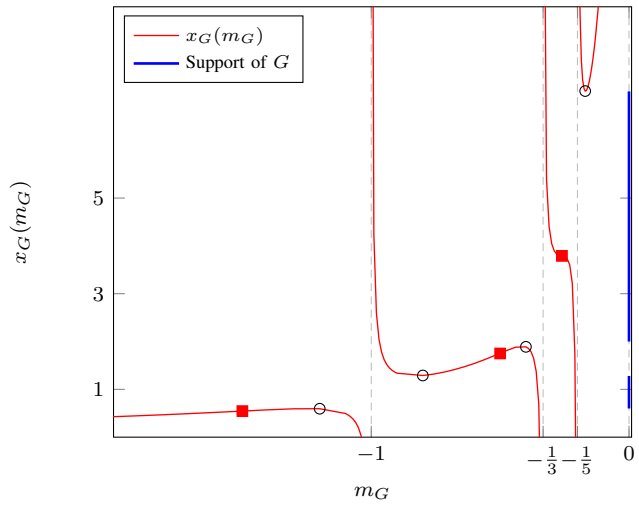

Fig. 8. $x_G(m_G)$ for m_G real, $\mathbf{P}$ diagonal composed of three evenly weighted masses in 1, 3 and 5. Local extrema are marked in circles, inflexion points are marked in squares.

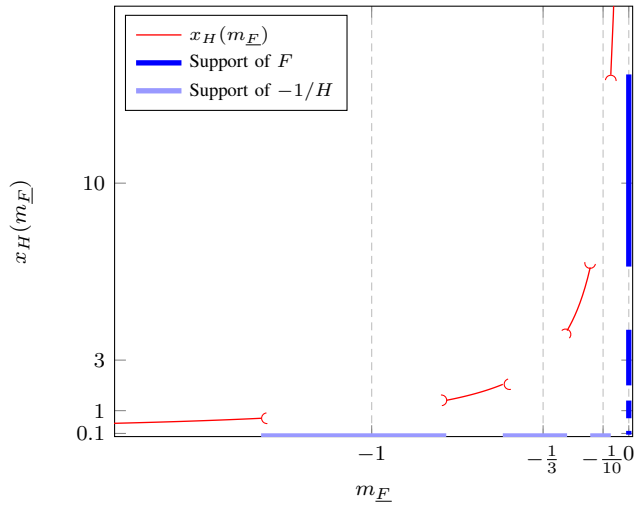

Fig. 9. $x_{\underline{F}}(m_{\underline{F}})$ for $m_{\underline{F}}$ real, $\sigma^2 = 0.1$, $c = c_0 = 10$, $\mathbf{P}$ diagonal composed of three evenly weighted masses in 1, 3 and 10. The support of F is read on the vertical axis.

support of T, as

$$x_A(m) = -\frac{1}{m} + \frac{1}{c}\int \frac{t}{1+tm} dT(t), \tag{33}$$

while, if x is outside the support of A, $m_A^\circ(x)$ is the only solution m of $x_A(m) = x$ such that $x'_A(m) > 0$. Moreover, if for some $m \in \mathbb{R}$ such that $-1/m$ is outside the support of T, $x'_A(m) > 0$, then $x_A(m)$ is outside the support of A.

The immediate corollary of Proposition 2 is that the complementary of the support $\mathrm{Supp}(A)$ of A is the set $\{x_A(m)\}$ for $-1/m$ outside the support of T such that $x'_A(m) > 0$,

$$\mathrm{Supp}(A) = \mathbb{R} \setminus \{x \mid \exists m \in \mathbb{R}, x = x_A(m), x'_A(m) > 0\}. \tag{34}$$

1) Support of G: First consider the matrix $\mathbf{HPH}^{\mathsf{H}}$, and let the function $x_G(m_G)$ be defined, for scalars $m_G \in \mathbb{R}^* \setminus \{-1/P_1, \ldots, -1/P_K\}$, by

$$x_G(m_G) = -\frac{1}{m_G} + \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r}{1+P_r m_G}. \tag{35}$$

The function $x_G(m_G)$ is depicted in Figures 7 and 8, for the cases where $c_0 = 10$, $c_1 = c_2 = c_3$ and (P_1, P_2, P_3) equal respectively $(1, 3, 10)$ and $(1, 3, 5)$. As expected by Proposition 2, $x_G(m_G)$ is increasing for m_G such that $x_G(m_G)$ is outside the support of G. Note now that the function x_G presents asymptotes in the positions $-1/P_1, \ldots, -1/P_K$,

$$\lim_{m_G \downarrow (-1/P_i)} x_G(m_G) = \infty \tag{36}$$

$$\lim_{m_G \uparrow (-1/P_i)} x_G(m_G) = -\infty, \tag{37}$$

and that $x_G(m_G) \to 0^+$ as $m_G \to -\infty$. Note also that, on its restriction to the set where it is non-decreasing, x_G is increasing.[2] To prove this, let m_G and $m_G^\star$ be two distinct points such that $x_G(m_G) > 0$ and $x_G(m_G^\star) > 0$, and $m_G^\star < m_G < 0$, we indeed have,[3]

$$x_G(m_G) - x_G(m_G^\star) = \frac{m_G - m_G^\star}{m_G m_G^\star} \times \left[1 - \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G})(P_r + \frac{1}{m_G^\star})}\right]. \tag{38}$$

Noticing that, for $P_i > 0$,

$$0 < \left(\frac{P_i}{P_i + \frac{1}{m_G}} - \frac{P_i}{P_i + \frac{1}{m_G^\star}}\right)^2 \tag{39}$$

$$= \frac{P_i^2}{(P_i + \frac{1}{m_G})^2} + \frac{P_i^2}{(P_i + \frac{1}{m_G^\star})^2} - 2\frac{P_i^2}{(P_i + \frac{1}{m_G})(P_i + \frac{1}{m_G^\star})}, \tag{40}$$

we have

$$\left(1 - \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G})^2}\right) + \left(1 - \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G^\star})^2}\right) < 2 - 2\sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G})(P_r + \frac{1}{m_G^\star})}. \tag{41}$$

Since we also have

$$x'_G(m_G) = \frac{1}{m_G^2}\left[1 - \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G})^2}\right] \geq 0 \tag{42}$$

$$x'_G(m_G^\star) = \frac{1}{(m_G^\star)^2}\left[1 - \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^2}{(P_r + \frac{1}{m_G^\star})^2}\right] \geq 0, \tag{43}$$

we conclude that the term in brackets in (38) is positive and then that $x_G(m_G) - x_G(m_G^\star) > 0$. Hence x_G is increasing on its restriction to the set where it is non-decreasing.

Notice also that x_G, both in Figures 7 and 8, has exactly one inflexion point on each open set $(-1/P_{i-1}, -1/P_i)$, for

[2] we say here that a function $f(x)$ is increasing if $x < x^\star \Rightarrow f(x) - f(x^\star) > 0$; if $x < x^\star \Rightarrow f(x) - f(x^\star) \geq 0$, we say that $f(x)$ is non-decreasing.

[3] this proof is borrowed from the proof of [15], with different notations.

$i \in \{1, \dots, K\}$, with convention $P_0 = 0+$. This is proven by noticing that $x''_G(m_G) = 0$ is equivalent to

$$\sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^3 m_G^3}{(1 + P_r m_G)^3} - 1 = 0. \tag{44}$$

Now, the left-hand side of (44) has derivative along m_G,

$$3 \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r^3 m_G^2}{(1 + P_r m_G)^4}, \tag{45}$$

which is always positive. Notice that the left-hand side of (44) has asymptotes for $m_G = -1/P_i$ for all $i \in \{1, \dots, K\}$, and has limits 0 as $m_G \to 0$ and $1/c_0 - 1$ as $m_G \to -\infty$. If $c_0 > 1$, Equation (44) (and then $x''_G(m_G) = 0$) therefore has a unique solution in $(-1/P_{i-1}, -1/P_i)$ for all $i \in \{1, \dots, K\}$. When x_G is increasing somewhere on $(-1/P_{i-1}, -1/P_i)$, the inflexion point, i.e., the solution to $x''_G(m_G) = 0$, in $(-1/P_{i-1}, -1/P_i)$ is necessarily found in the region where x_G increases. If $c_0 \leq 1$, the leftmost inflexion point may not exist.

From the discussion above and Proposition 2, it is clear that the support of G is divided into $K_G \leq K$ compact subsets $[x_{G,i}^-, x_{G,i}^+]$, $i \in \{1, \dots, K_G\}$. Also, if $c_0 > 1$, G has an additional mass in 0 of probability $G(0) - G(0-) = (c_0 - 1)/c_0$; this mass will not be counted as a cluster in G. Observe that every P_i can be uniquely mapped to a corresponding subset $[x_{G,j}^-, x_{G,j}^+]$ in the following fashion. The power P_1 is mapped onto the first cluster in G; we then have $1_G = 1$. Then the power P_2 is either mapped onto the second cluster in G if x_G increases in the subset $(-1/P_1, -1/P_2)$, which is equivalent to saying that $x'_G(m_{G,2}) > 0$ for $m_{G,2}$ the only solution to $x''_G(m_G) = 0$ in $(-1/P_1, -1/P_2)$; in this case, we have $2_G = 2$ and the clusters associated to P_1 and P_2 in G are distinct. Otherwise, if $x'_G(m_{G,2}) \leq 0$, P_2 is mapped onto the first cluster in F; in this case, $2_G = 1$. The latter scenario visually corresponds to the case when P_1 and P_2 engender "overlapping clusters". More generally, P_j, $j \in \{1, \dots, K\}$, is uniquely mapped onto the cluster j_G such that

$$j_G = \#\left\{i \leq j \mid \min[x'_G(m_{G,i}), x'_G(m_{G,i+1})] > 0\right\}, \tag{46}$$

with convention $m_{G,K+1} = 0$, which is exactly

$$j_G = \#\left\{i \leq j \mid i \text{ satisfies Assumption 1}\right\}, \tag{47}$$

when $c_0 > 1$. If $c_0 \leq 1$, $m_{G,1}$, the zero of x''_G in $(-\infty, -1/P_1)$ may not exist. If $c_0 < 1$, we claim that P_1 cannot be evaluated (as was already observed in Remark 2). The special case when $c_0 = 1$ would require a restatement of Assumption 1 to handle the special case of P_1; this will however not be done, as it will turn out that Assumption 2 is violated for P_1 if $\sigma^2 > 0$, which we assume.

In the particular case of the power P_k of interest in Theorem 2, because of Assumption 1, $x'_G(m_{G,k}) > 0$. Therefore the index k_G of the cluster associated to P_k in G satisfies $k_G = (k-1)_G + 1$ (with convention $0_G = 0$). Also, from Assumption 1, $x'_G(m_{G,k+1}) > 0$. Therefore $(k+1)_G = k_G + 1$. In that case, we have that P_k is the only power mapped to cluster k_G in G, and then we have the required cluster separability condition.

2) Support of F: We now proceed to the study of F, the almost sure limit spectrum distribution of $\mathbf{B}_N$. In the same way as in the previous section, we have that the support of $\underline{F}$ is fully determined by the function $x_{\underline{F}}(m_{\underline{F}})$, defined for $m_{\underline{F}}$ real, such that $-1/m_{\underline{F}}$ lies outside the support of H, by

$$x_{\underline{F}}(m_{\underline{F}}) = -\frac{1}{m_{\underline{F}}} + \frac{1 + c_0}{c c_0} \int \frac{t}{1 + t m_{\underline{F}}} dH(t). \tag{48}$$

Figure 9 depicts the function $x_{\underline{F}}$ in the case of Figure 2, i.e., $K = 3$, $P_1 = 1, P_2 = 3, P_3 = 10$, $c_1 = c_2 = c_3$, $c_0 = 10$, $c = 10$, $\sigma^2 = 0.1$. Figure 9 has the peculiar behaviour that it does not have asymptotes as in Figure 7 where the population eigenvalue distribution was discrete. As a consequence, our previous derivations cannot be straightforwardly adapted to derive the spectrum separability condition. If $c_0 > 1$, note also, although it is not appearing in the abscissa range of Figure 9, that there exist asymptotes in the position $m_{\underline{F}} = -1/\sigma^2$. This is due to the fact that $G(0) - G(0-) > 0$, and therefore $H(\sigma^2) - H(\sigma^2-) > 0$. We assume $c_0 > 1$ until further notice.

From Proposition 2, the support of $\underline{F}$ is complementary to the set of real nonnegative x such that $x = x_{\underline{F}}(m_{\underline{F}})$ and $x'_{\underline{F}}(m_{\underline{F}}) > 0$ for a certain real $m_{\underline{F}}$, with $x'_{\underline{F}}(m_{\underline{F}})$ given by

$$x'_{\underline{F}}(m_{\underline{F}}) = \frac{1}{{m_{\underline{F}}}^2} - \frac{1 + c_0}{c c_0} \int \frac{t^2}{(1 + t m_{\underline{F}})^2} dH(t). \tag{49}$$

Reminding that $H(t) = \frac{c_0}{c_0+1} G(t - \sigma^2) + \frac{1}{1+c_0} \delta(t)$, this can be rewritten

$$x'_{\underline{F}}(m_{\underline{F}}) = \frac{1}{{m_{\underline{F}}}^2} - \frac{1}{c} \int \frac{t^2}{(1 + t m_{\underline{F}})^2} dG(t - \sigma^2). \tag{50}$$

It is still true that $x_{\underline{F}}(m_{\underline{F}})$, restricted to the set of $m_{\underline{F}}$ where $x'_{\underline{F}}(m_{\underline{F}}) \geq 0$, is increasing. As a consequence, it is still true also that each cluster of H can be mapped to a unique cluster in $\underline{F}$. It is then possible to iteratively map the power P_k onto cluster k_G in G, as previously described, and to further map cluster k_G in G (which is also cluster k_G in H) onto a unique cluster k_F in $\underline{F}$ (or equivalently in F).

Therefore, a necessary and sufficient condition for the separability of the cluster associated to P_k in $\underline{F}$ reads

Assumption 3: There exist two distinct real values $m_{\underline{F},k_G}^{(l)} < m_{\underline{F},k_G}^{(r)}$ such that

1) $x'_{\underline{F}}(m_{\underline{F},k_G}^{(l)}) > 0$, $x'_{\underline{F}}(m_{\underline{F},k_G}^{(r)}) > 0$
2) there exist $m_{G,k}^{(l)}, m_{G,k}^{(r)} \in \mathbb{R}$ such that $x_G(m_{G,k}^{(l)}) = -1/m_{\underline{F},k_G}^{(l)} - \sigma^2$ and $x_G(m_{G,k}^{(r)}) = -1/m_{\underline{F},k_G}^{(r)} - \sigma^2$ that satisfy
   a) $x'_G(m_{G,k}^{(l)}) > 0$, $x'_G(m_{G,k}^{(r)}) > 0$,
   b) and

$$P_{k-1} < -\frac{1}{m_{G,k}^{(l)}} < P_k < -\frac{1}{m_{G,k}^{(r)}} < P_{k+1} \tag{51}$$

with the convention $P_0 = 0+$, $P_{K+1} = \infty$.

Assumption 3 states (i) that cluster k_G in G is distinct from clusters $(k-1)_G$ and $(k+1)_G$ (Item 2b); this is another way of stating Assumption 1, and (ii) that the points $m_{\underline{F},k_G}^{(l)} \triangleq -1/(x_G(m_{G,k_G}^{(l)}) + \sigma^2)$ and $m_{\underline{F},k_G}^{(r)} \triangleq -1/(x_G(m_{G,k_G}^{(r)}) + \sigma^2)$ (which lie on either side of cluster k_G in H) have respective

images $x_{k_F}^{(l)} \triangleq x_{\underline{F}}(m_{\underline{F},k_G}^{(l)})$ and $x_{k_F}^{(r)} \triangleq x_{\underline{F}}(m_{\underline{F},k_G}^{(r)})$ by $x_{\underline{F}}$, such that $x'_{\underline{F}}(m_{\underline{F},k_G}^{(l)}) > 0$ and $x'_{\underline{F}}(m_{\underline{F},k_G}^{(r)}) > 0$, i.e., $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$ lie outside the support of $\underline{F}$, on either side of cluster k_F.

However, Assumption 3, be it a necessary and sufficient condition for the separability of cluster k_F, is difficult to exploit in practice. Indeed, it is not satisfactory to require the verification of the existence of such $m_{\underline{F},k_G}^{(l)}$ and $m_{\underline{F},k_G}^{(r)}$. More importantly, the computation of $x_{\underline{F}}$ requires to know H, which is only fully accessible through the non-convenient inverse Stieltjes transform formula

$$H(x) = \frac{1}{\pi} \lim_{y \to 0} \int_{-\infty}^{x} m_H(t+iy)dt. \tag{52}$$

Instead of Assumption 3, we derive here a sufficient condition for cluster separability in $\underline{F}$. Notice from the clustering of G into K_G clusters plus a mass at zero that (50) becomes

$$x'_{\underline{F}}(m_{\underline{F}}) = \frac{1}{{m_{\underline{F}}}^2} - \frac{1}{c} \sum_{r=1}^{K_G} \int_{x_{G,r}^-}^{x_{G,r}^+} \frac{t^2}{(1+tm_{\underline{F}})^2} dG(t-\sigma^2) - \frac{c_0 - 1}{cc_0} \frac{\sigma^4}{(1+\sigma^2 m_{\underline{F}})^2}, \tag{53}$$

where we remind that $[x_{G,i}^-, x_{G,i}^+]$ is the support of cluster i in G, i.e., $x_{G,1}^-, x_{G,1}^+, \ldots, x_{G,K_G}^-, x_{G,K_G}^+$ are the images by x_G of the $2K_G$ real solutions to $x'_G(m_G) = 0$.

Observe now that the function $-t^2/(1+tm_{\underline{F}})^2$, found in the integrals of (53), has derivative along t

$$\left(-\frac{t^2}{(1+tm_{\underline{F}})^2} \right)' = -\frac{2t}{(1+tm_{\underline{F}})^4}(1+tm_{\underline{F}}) \tag{54}$$

and is therefore strictly increasing when $m_{\underline{F}} < -1/t$ and strictly decreasing when $m_{\underline{F}} > -1/t$. For $m_{\underline{F}} \in (-1/(x_{G,i}^+ + \sigma^2), -1/x_{G,i+1}^- + \sigma^2)$, we then have the inequality

$$x'_{\underline{F}}(m_{\underline{F}}) \geq \frac{1}{{m_{\underline{F}}}^2} - \frac{1}{c} \left(\sum_{r=1}^{i} \frac{(x_{G,r}^+ + \sigma^2)^2}{(1+(x_{G,r}^+ + \sigma^2)m_{\underline{F}})^2} + \sum_{r=i+1}^{K_G} \frac{(x_{G,r}^- + \sigma^2)^2}{(1+(x_{G,r}^- + \sigma^2)m_{\underline{F}})^2} + \frac{c_0 - 1}{c_0} \frac{\sigma^4}{(1+\sigma^2 m_{\underline{F}})^2} \right). \tag{55}$$

Denote $f_i(m_{\underline{F}})$ the right-hand side of (55). Through the inequality (55), we then fall back on a finite sum expression as in the previous study of the support of G. In that case, we can exhibit a sufficient condition to ensure the separability of cluster k_F from the neighboring clusters. Specifically, we only need to verify that $f_{k_G-1}(m_{\underline{F},k_G}) > 0$, with $m_{\underline{F},k_G}$ the single solution to $f'_{k_G-1}(m_{\underline{F}}) = 0$ in the set $(-1/(x_{G,k_G-1}^+ + \sigma^2), -1/(x_{G,k_G}^- + \sigma^2))$, and $f_{k_G}(m_{\underline{F},k_G+1}) > 0$, with $m_{\underline{F},k_G+1}$ the unique solution to $f'_{k_G}(m_{\underline{F}}) = 0$ in the set $(-1/(x_{G,k_G}^+ + \sigma^2), -1/(x_{G,k_G+1}^- + \sigma^2))$. This is exactly what Assumption 2 states.

Remember now that we assumed in this section $c_0 > 1$. If $c_0 \leq 1$, then 0 is in the support of H and therefore the leftmost cluster in F, i.e., that attached to σ^2, is necessarily merged with that of P_1. This already discards the possibility of spectrum separation for P_1 and therefore P_1 cannot be estimated. It is therefore not necessary to update Assumption 1 for the particular case of P_1, when $c_0 = 1$.

Therefore, Assumptions 1 and 2 ensure that $(k-1)_F < k_F < (k+1)_F$, $k_F \neq 1$, and there exists a constructive way to derive the mapping $k \mapsto k_F$. We are now in position to determine the contour $\mathcal{C}_k$.

B. Determination of $\mathcal{C}_k$

From Assumption 2 and Proposition 2, there exist $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$ outside the support of F, on either side of cluster k_F, such that $m_{\underline{F}}(z)$ has limits $m_{\underline{F},k_G}^{(l)} \triangleq {m_{\underline{F}}}^\circ(x_{k_F}^{(l)})$ and $m_{\underline{F},k_G}^{(r)} \triangleq {m_{\underline{F}}}^\circ(x_{k_F}^{(r)})$, as $z \to x_{k_F}^{(l)}$ and $z \to x_{k_F}^{(r)}$, respectively, with ${m_{\underline{F}}}^\circ$ the analytic extension of $m_{\underline{F}}$ in the points $x_{k_F}^{(l)} \in \mathbb{R}$ and $x_{k_F}^{(r)} \in \mathbb{R}$. These limits $m_{\underline{F},k_G}^{(l)}$ and $m_{\underline{F},k_G}^{(r)}$ are on either side of cluster k_G in the support of $-1/H$, and therefore $-1/m_{\underline{F},k_G}^{(l)} - \sigma^2$ and $-1/m_{\underline{F},k_G}^{(l)} - \sigma^2$ are on either side of cluster k_G in the support of G.

Consider any continuously differentiable complex path $\Gamma_{F,k}$ with endpoints $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$, and interior points of positive imaginary part. We define the contour $\mathcal{C}_{F,k}$ as the union of $\Gamma_{F,k}$ oriented from $x_{k_F}^{(l)}$ to $x_{k_F}^{(r)}$ and its complex conjugate $\Gamma_{F,k}^*$ oriented backwards from $x_{k_F}^{(r)}$ to $x_{k_F}^{(l)}$. The contour $\mathcal{C}_{F,k}$ is clearly continuous and piecewise continuously differentiable. Also, the support of cluster k_F in $\underline{F}$ is completely inside $\mathcal{C}_{F,k}$, while the supports of the neighboring clusters are away from $\mathcal{C}_{F,k}$. The support of cluster k_G in H is then inside $-1/m_{\underline{F}}(\mathcal{C}_{F,k})$,[4] and therefore the support of cluster k_G in G is inside $\mathcal{C}_{G,k} \triangleq -1/m_{\underline{F}}(\mathcal{C}_{F,k}) - \sigma^2$. Since $m_{\underline{F}}$ is continuously differentiable on $\mathbb{C} \setminus \mathbb{R}$ (it is in fact holomorphic there [19]) and has limits in $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$, $\mathcal{C}_{G,k}$ is also continuous and piecewise continuously differentiable. Going one last step in this process, we finally have that P_k is inside the contour $\mathcal{C}_k \triangleq -1/m_G(\mathcal{C}_{G,k})$, while P_i, for all $i \neq k$, is outside $\mathcal{C}_k$. Since m_G is also holomorphic on $\mathbb{C} \setminus \mathbb{R}$ and has limits in $-1/{m_{\underline{F}}}^\circ(x_{k_F}^{(l)}) - \sigma^2$ and $-1/{m_{\underline{F}}}^\circ(x_{k_F}^{(r)}) - \sigma^2$, $\mathcal{C}_k$ is a continuous and piecewise continuously differentiable complex path, which is sufficient to perform complex integration [18].

The contours $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ originating from circular integration contours $\mathcal{C}_{F,k}$ of diameter $[x_{k_F}^{(l)}, x_{k_F}^{(r)}]$, $k \in \{1,2,3\}$, for the case of Figure 2, are depicted in Figure 10. The points $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$ for $k_F \in \{1,2,3\}$ are taken to be $x_{k_F}^{(l)} = x_{\underline{F}}(m_{\underline{F},k_G})$, $x_{k_F}^{(r)} = x_{\underline{F}}(m_{\underline{F},k_G+1})$, with $m_{\underline{F},i}$ the real root of $f_i'(m_{\underline{F}}) = 0$ in $(-1/(x_{G,i-1}^+ + \sigma^2), -1/(x_{G,i}^- + \sigma^2))$ when $i \in \{1,2,3\}$, and we take the convention $m_{G,4} = -1/(15+\sigma^2)$.

Recall now that P_k was defined as

$$P_k = c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{P_r - \omega} d\omega. \tag{56}$$

[4] we slightly abuse notations here and should instead say that the support of cluster k_G in H is inside the contour described by the image by $-1/m_{\underline{F}}$ of the restriction to $\mathbb{C}^+$ and $\mathbb{C}^-$ of $\mathcal{C}_{F,k}$, continuously extended to $\mathbb{R}$ in the points $-1/m_{\underline{F},k_G}^{(l)}$ and $-1/m_{\underline{F},k_G}^{(r)}$.

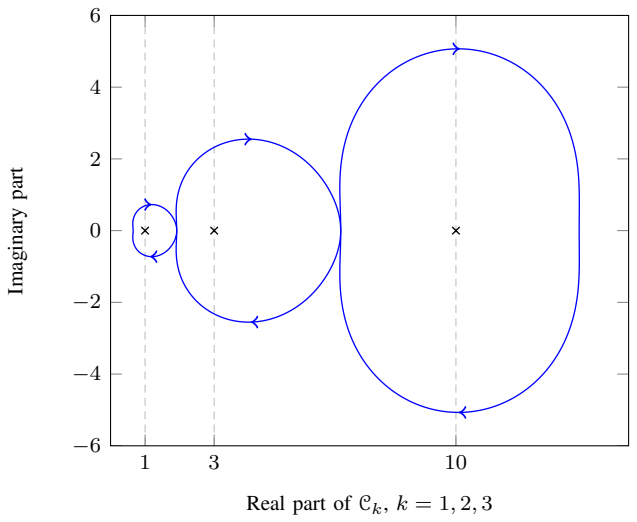

Fig. 10. Integration contours $\mathcal{C}_{F,1}$, $\mathcal{C}_{F,2}$ and $C_{F,3}$, for $c = 10$, $c_0 = 10$, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$.

With the variable change $\omega = -1/m_G(t)$, this becomes

$$P_k = \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{G,k}} \left(m_G(t) \left[-\frac{1}{m_G(t)} + \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r}{1 + P_r m_G(t)} \right] + \frac{c_0 - 1}{c_0} \right) \frac{m_G'(t)}{m_G(t)^2} dt. \quad (57)$$

From Equation (12), this simplifies into

$$P_k = \frac{c_k}{c_0} \frac{1}{2\pi i} \oint_{\mathcal{C}_{G,k}} (c_0 t m_G(t) + (c_0 - 1)) \frac{m_G'(t)}{m_G(t)^2} dt. \quad (58)$$

Using (16) and proceeding to the further change of variable $t = -1/m_{\underline{F}}(z) - \sigma^2$, (58) becomes

$$P_k = \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{F,k}} \left[\left(\frac{1}{m_{\underline{F}}(z)} + \sigma^2 \right) z m_{\underline{F}}(z) m_F(z) + \frac{c_0 - 1}{c_0} \right] \times \frac{-m_{\underline{F}}(z) m_F(z) - z m_{\underline{F}}'(z) m_F(z) - z m_{\underline{F}}(z) m_F'(z)}{z^2 m_{\underline{F}}(z)^2 m_F(z)^2} dz \quad (59)$$

$$= \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{F,k}} \left[\left(1 + \sigma^2 m_{\underline{F}}(z)\right) + \frac{c_0 - 1}{c_0} \frac{1}{z m_F(z)} \right] \times \left[-\frac{1}{z m_{\underline{F}}(z)} - \frac{m_{\underline{F}}'(z)}{m_{\underline{F}}(z)^2} - \frac{m_F'(z)}{m_F(z) m_{\underline{F}}(z)} \right] dz. \quad (60)$$

This whole process of variable changes allowed us to describe P_k as a function of $m_F(z)$, the Stieltjes transform of the almost sure limiting spectral distribution of $\mathbf{B}_N$, as $N \to \infty$. It then remains to exhibit a relation between P_k and the empirical spectral distribution of $\mathbf{B}_N$ for finite N. This is to what the subsequent section is dedicated to.

C. Evaluation of $\hat{P}_k$

Let us now define $\hat{m}_F(z)$ and $\hat{m}_{\underline{F}}(z)$ as the Stieltjes transforms of the empirical eigenvalue distributions of $\mathbf{B}_N$ and $\underline{\mathbf{B}}_N$, respectively, i.e.,

$$\hat{m}_F(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} \quad (61)$$

and

$$\hat{m}_{\underline{F}}(z) = \frac{N}{M} \hat{m}_F(z) - \frac{M - N}{M} \frac{1}{z}. \quad (62)$$

Instead of going further with (59), define $\hat{P}_k$, the "empirical counterpart" of P_k, as

$$\hat{P}_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_{F,k}} \left[\frac{N}{n} \left(1 + \sigma^2 \hat{m}_{\underline{F}}(z)\right) + \frac{N - n}{n} \frac{1}{z \hat{m}_F(z)} \right] \times \left[-\frac{1}{z \hat{m}_{\underline{F}}(z)} - \frac{\hat{m}_{\underline{F}}'(z)}{\hat{m}_{\underline{F}}(z)^2} - \frac{\hat{m}_F'(z)}{\hat{m}_F(z) \hat{m}_{\underline{F}}(z)} \right] dz. \quad (63)$$

The integrand can then be expanded into nine terms, for which residue calculus [18] can easily be performed. Denote first $\eta_1, \ldots, \eta_N$ the N real roots of $\hat{m}_F(z) = 0$ and $\mu_1, \ldots, \mu_N$ the N real roots of $\hat{m}_{\underline{F}}(z) = 0$. We identify three sets of possible poles for the nine aforementioned terms: (i) the set $\{\lambda_1, \ldots, \lambda_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$, (ii) the set $\{\eta_1, \ldots, \eta_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$ and (iii) the set $\{\mu_1, \ldots, \mu_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$. For $M \neq N$, the full calculus leads to

$$\hat{P}_k = \frac{NM}{n_k(M - N)} \left[\sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \eta_i \le x_{k_F}^{(r)}}} \eta_i - \sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \mu_i \le x_{k_F}^{(r)}}} \mu_i \right] + \frac{N}{n_k} \left[\sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \eta_i \le x_{k_F}^{(r)}}} \sigma^2 - \sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \lambda_i \le x_{k_F}^{(r)}}} \sigma^2 \right] + \frac{N}{n_k} \left[\sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \mu_i \le x_{k_F}^{(r)}}} \sigma^2 - \sum_{\substack{1 \le i \le N \\ x_{k_F}^{(l)} \le \lambda_i \le x_{k_F}^{(r)}}} \sigma^2 \right]. \quad (64)$$

Details are given in Appendix A. Now, we know from Theorem 1 that $\hat{m}_F(z) \xrightarrow{\text{a.s.}} m_F(z)$ and $\hat{m}_{\underline{F}}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$ as $N \to \infty$. Observing that the integrand in (63) is uniformly bounded on the compact $\mathcal{C}_{F,k}$, the dominated convergence theorem [20] ensures $\hat{P}_k \xrightarrow{\text{a.s.}} P_k$.

To go further, we now need to determine which of $\lambda_1, \ldots, \lambda_N$, $\eta_1, \ldots, \eta_N$ and $\mu_1, \ldots, \mu_N$ lie inside $\mathcal{C}_{F,k}$. This requires a result of eigenvalue *exact separation* that extends the earlier results of [21], [17], as follows

Theorem 3: Let $\mathbf{B}_n = (1/n) \mathbf{T}_n^{\frac{1}{2}} \mathbf{X}_n \mathbf{X}_n^{\mathsf{H}} \mathbf{T}_n^{\frac{1}{2}} \in \mathbb{C}^{p \times p}$, where we assume the following conditions

1) $\mathbf{X}_n \in \mathbb{C}^{p \times n}$ has entries x_{ij}, $1 \le i \le p$, $1 \le j \le n$, extracted from a doubly infinite array $\{x_{ij}\}$ of independent variables, with zero mean and unit variance.
2) There exist K and a random variable X with finite fourth order moment such that, for any $x > 0$,

$$\frac{1}{n_1 n_2} \sum_{i \le n_1, j \le n_2} P(|x_{ij}| > x) \le K P(|X| > x) \quad (65)$$

for any n_1, n_2.

3) There is a positive function $\psi(x) \uparrow \infty$ as $x \to \infty$, and $M > 0$, such that
$$\max_{ij} \mathrm{E}|x_{ij}^2|\psi(|x_{ij}|) \leq M. \quad (66)$$
4) $p = p(n)$ with $c_n = p/n \to c > 0$ as $n \to \infty$.
5) For each n, $\mathbf{T}_n \in \mathbb{C}^{p\times p}$ is Hermitian nonnegative definite, independent of $\{x_{ij}\}$, satisfying $H_n \triangleq F^{\mathbf{T}_n} \Rightarrow H$, H a nonrandom probability distribution function, almost surely. $\mathbf{T}_n^{\frac{1}{2}}$ is any Hermitian square root of $\mathbf{T}_n$.
6) The spectral norm $\|\mathbf{T}_n\|$ of $\mathbf{T}_n$ is uniformly bounded in n almost surely.
7) Let $a, b > 0$, nonrandom, be such that, with probability one, $[a, b]$ lies in an open interval outside the support of F^{c_n,H_n} for all large n, with $F^{y,G}$ defined to be the almost sure l.s.d. of $(1/n)\mathbf{X}_n^{\mathsf{H}}\mathbf{T}_n\mathbf{X}_n$ when $H = G$ and $c = y$.

Denote $\lambda_1^{\mathbf{Y}} \geq \ldots \geq \lambda_p^{\mathbf{Y}}$ the ordered eigenvalues of the Hermitian matrix $\mathbf{Y} \in \mathbb{C}^{p\times p}$. Then, we have that

1) $P(\text{no eigenvalue of } \mathbf{B}_n \text{ lies in } [a, b] \text{ for all large } n) = 1$.
2) If $c(1 - H(0)) > 1$, then x_0, the smallest value in the support of $F^{c,H}$, is positive, and with probability one, $\lambda_n^{\mathbf{B}_n} \to x_0$ as $n \to \infty$.
3) If $c(1 - H(0)) \leq 1$, or $c(1 - H(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then $m_{F^{c,H}}(a) < m_{F^{c,H}}(b) < 0$. With probability one, there exists, for all n large, an index $i_n \geq 0$ such that $\lambda_{i_n}^{\mathbf{T}_n} > -1/m_{F^{c,H}}(b)$ and $\lambda_{i_n+1}^{\mathbf{T}_n} > -1/m_{F^{c,H}}(a)$ and we have
$$P(\lambda_{i_n}^{\mathbf{B}_n} > b \text{ and } \lambda_{i_n+1}^{\mathbf{B}_n} < a \text{ for all large } n) = 1. \quad (67)$$

Theorem 3 is proven in Appendix B.

To apply Theorem 3 to $\underline{\mathbf{B}}_N$ in our scenario, we need to ensure all assumptions are met. Only Items 2-6 need particular attention. In our scenario, the matrix $\mathbf{X}_n$ of Theorem 3 is $\left(\begin{smallmatrix}\mathbf{X}\\ \mathbf{W}\end{smallmatrix}\right)$, while $\mathbf{T}_n$ is $\mathbf{T} \triangleq \left(\begin{smallmatrix}\mathbf{H}\mathbf{P}\mathbf{H}^{\mathsf{H}}+\sigma^2\mathbf{I}_N & 0\\ 0 & 0\end{smallmatrix}\right)$. The latter has been proven to have almost sure l.s.d. H, so that Item 5 is verified. Also, from the result of [21] upon which Theorem 3 is based, there exists a subset of probability one in the probability space that engenders the $\mathbf{T}$ over which, for n large enough, $\mathbf{T}$ has no eigenvalues in any closed set strictly outside the support of H; this ensures Item 6. Now, from construction, $\mathbf{X}$ and $\mathbf{W}$ have independent entries of zero mean, unit variance, fourth order moment and are composed of at most $K + 1$ distinct distributions, irrespectively of M. Denote $X_1, \ldots, X_d$, $d \leq K + 1$, d random variables distributed as those distinct distributions. Letting $X = |X_1| + \ldots + |X_d|$, we have that
$$\frac{1}{n_1 n_2} \sum_{i\leq n_1, j\leq n_2} P(|z_{ij}| > x) \leq P\left(\sum_{i=1}^{d} |X_i| > x\right) \quad (68)$$
$$= P(|X| > x), \quad (69)$$
where z_{ij} is the $(i, j)^{th}$ entry of $\left(\begin{smallmatrix}\mathbf{X}\\ \mathbf{W}\end{smallmatrix}\right)$. Since all X_i have finite order four moments, so does X and Item 2 is verified. From the same argument, Item 3 follows with $\phi(x) = x^2$. Theorem 3 can then be applied to $\underline{\mathbf{B}}_N$.

The corollary of Theorem 3 applied to $\underline{\mathbf{B}}_N$ is that, with probability one, for N sufficiently large, there will be no eigenvalue of $\mathbf{B}_N$ (or $\underline{\mathbf{B}}_N$) outside the support of F, and the number of eigenvalues inside cluster k_F is exactly n_k. Since $\mathcal{C}_{F,k}$ encloses cluster k_F and is away from the other clusters, $\{\lambda_1, \ldots, \lambda_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}] = \{\lambda_i, i \in \mathcal{N}_k\}$ almost surely, for all large N. Also, for any $i \in \{1, \ldots, N\}$, it is easy to see from (61) that $\hat{m}_F(z) \to \infty$ when $z \uparrow \lambda_i$ and $\hat{m}_F(z) \to -\infty$ when $z \downarrow \lambda_i$. Therefore $\hat{m}_F(z) = 0$ has at least one solution in each interval $(\lambda_{i-1}, \lambda_i)$, with $\lambda_0 = 0$, hence $\mu_1 < \lambda_1 < \mu_2 < \ldots < \mu_N < \lambda_N$. This implies that, if k_0 is the index such that $\mathcal{C}_{F,k}$ contains exactly $\lambda_{k_0}, \ldots, \lambda_{k_0+(n_k-1)}$, then $\mathcal{C}_{F,k}$ also contains $\{\mu_{k_0+1}, \ldots, \mu_{k_0+(n_k-1)}\}$. The same result holds for $\eta_{k_0+1}, \ldots, \eta_{k_0+(n_k-1)}$. When the indexes exist, due to cluster separability, η_{k_0-1} and μ_{k_0-1} belong, for N large, to cluster $k_F - 1$. We are then left with determining whether μ_{k_0} and η_{k_0} are asymptotically found inside $\mathcal{C}_{F,k}$.

For this, we use the same approach as in [15], by noticing that, since 0 is not included in $\mathcal{C}_k$, one has
$$\frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{\omega} d\omega = 0. \quad (70)$$

Performing the same changes of variables as above, we have that
$$\oint_{\mathcal{C}_{F,k}} \frac{-m_{\underline{F}}(z)m_F(z) - z m_{\underline{F}}'(z)m_F(z) - z m_{\underline{F}}(z)m_F'(z)}{z^2 m_{\underline{F}}(z)^2 m_F(z)^2} dz = 0. \quad (71)$$

For N large, the dominated convergence theorem ensures again that the left-hand side of the (71) is close to
$$\oint_{\mathcal{C}_{F,k}} \frac{-\hat{m}_{\underline{F}}(z)\hat{m}_F(z) - z\hat{m}_{\underline{F}}'(z)\hat{m}_F(z) - z\hat{m}_{\underline{F}}(z)\hat{m}_F'(z)}{z^2 \hat{m}_{\underline{F}}(z)^2 \hat{m}_F(z)^2} dz. \quad (72)$$

Residue calculus of (72) then leads to
$$\left[\sum_{\substack{1\leq i\leq N\\ \lambda_i\in[x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 2 - \sum_{\substack{1\leq i\leq N\\ \eta_i\in[x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 1 - \sum_{\substack{1\leq i\leq N\\ \mu_i\in[x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 1\right] \xrightarrow{\text{a.s.}} 0. \quad (73)$$

Since the cardinalities of $\{i, \eta_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]\}$ and $\{i, \mu_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]\}$ are at most n_k, (73) is satisfied only if both cardinalities equal n_k in the limit. As a consequence, $\mu_{k_0} \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$ and $\eta_{k_0} \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$. For N large, $N \neq M$, this allows us to simplify (64) into
$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{\substack{1\leq i\leq N\\ \lambda_i\in\mathcal{N}_k}} (\eta_i - \mu_i) \quad (74)$$
with probability one. The same reasoning holds for $M = N$. This is our final relation.

It now remains to show that the η_i and the μ_i are the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ respectively. For this, we need the following lemma,

Lemma 1: Let $\mathbf{A} \in \mathbb{R}^{N\times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$, and let $\mathbf{y} \in \mathbb{R}^N$. Then the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathsf{H}}$ are the N real solutions of the following equation in x,
$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1. \quad (75)$$

Proof: Let λ be an eigenvalue of $\mathbf{A}-\mathbf{y}\mathbf{y}^{\mathsf{H}}$. For a certain non-zero vector $\mathbf{x}\in\mathbb{C}^N$, we then have the equivalent relations

$$(\mathbf{A}-\mathbf{y}\mathbf{y}^{\mathsf{H}})\mathbf{x}=\lambda\mathbf{x}, \tag{76}$$

$$(\mathbf{A}-\lambda\mathbf{I}_N)\mathbf{x}=\mathbf{y}^{\mathsf{H}}\mathbf{x}\mathbf{y}, \tag{77}$$

$$\mathbf{x}=\mathbf{y}^{\mathsf{H}}\mathbf{x}(\mathbf{A}-\lambda\mathbf{I}_N)^{-1}\mathbf{y}, \tag{78}$$

$$\mathbf{y}^{\mathsf{H}}\mathbf{x}=\mathbf{y}^{\mathsf{H}}\mathbf{x}\mathbf{y}^{\mathsf{H}}(\mathbf{A}-\lambda\mathbf{I}_N)^{-1}\mathbf{y}, \tag{79}$$

$$1=\mathbf{y}^{\mathsf{H}}(\mathbf{A}-\lambda\mathbf{I}_N)^{-1}\mathbf{y}. \tag{80}$$

Since $\mathbf{A}$ is diagonal, denoting $\mathbf{e}_i\in\mathbb{C}^N$ the vector such that $e_{i,j}=\delta_i^j$, we finally have

$$\sum_{i=1}^{N}\frac{(\mathbf{y}^{\mathsf{H}}\mathbf{e}_i)^2}{\lambda_i-\lambda}=1. \tag{81}$$

■

Applying Lemma 1 to $\mathbf{A}=\operatorname{diag}\boldsymbol{\lambda}$ and $\mathbf{y}=\sqrt{\frac{1}{N}\boldsymbol{\lambda}}$, we find that the eigenvalues of $\operatorname{diag}(\boldsymbol{\lambda})-\frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ are the solutions of

$$\sum_{i=1}^{N}\frac{\frac{1}{N}\lambda_i}{\lambda_i-x}=1, \tag{82}$$

which is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i-x}=0, \tag{83}$$

whose solutions are by definition $\eta_1,\ldots,\eta_N$. The same argument applies similarly to $\mu_1,\ldots,\mu_N$. Incidentally, this remark was already noticed in [22].

We end this section by a short discussion on the consequences of Theorem 2.

D. Discussion

Theorem 2 states that, under spectrum separability condition for all P_k, $k\in\{1,\ldots,K\}$, when $n_1,\ldots,n_K$ are known *a priori* to the receiver, then $\hat{P}_1,\ldots,\hat{P}_K$ are consistent estimators for $P_1,\ldots,P_K$. Now, in practice, it is rare that $n_1,\ldots,n_K$ and even K are *a priori* known to the receiver. However, if separability is assumed, then one can estimate simultaneously $K,n_1,\ldots,n_K$ and $P_1,\ldots,P_K$. This is performed by (i) determining the clusters of the empirical eigenvalues of $\mathbf{B}_N$, which determines K, (ii) counting the number of eigenvalues in each cluster to determine the multiplicities $n_1,\ldots,n_K$ and (iii) evaluating $\hat{P}_1,\ldots,\hat{P}_K$ from Theorem 2.

However, step (i) may not be obvious. In particular, when the total number n of transmit antennas is small, when the typical cluster size is large or when the inter-cluster spacing is small, it is non-trivial to determine what eigenvalues form a cluster. To solve this critical issue, studies are being currently carried out that aim to determine second order statistics of $F^{\mathbf{B}_N}$. Thanks to second order statistics on $F^{\mathbf{B}_N}$, it will be possible to design estimators of $P_1,\ldots,P_K$ that take into account the probability of $\mathbf{B}_N$ being an appropriate model for the estimated $\hat{P}_1,\ldots,\hat{P}_{\hat{K}}$ for every hypothesis $\hat{K}$ for the number of transmit source and every hypothesis $(\hat{n}_1,\ldots,\hat{n}_{\hat{K}})$ for the number of antennas for each of these sources. We hereafter provide an alternative *ad-hoc* technique to partially solve the problem of determining K and $n_1,\ldots,n_K$ based on Theorems 1 and 2.

In the following, we assume for readability that we know the number K of transmit sources (taken large enough to cover all possible hypotheses), some having possibly 0 transmit antennas. The approach consists in the following steps:

1) we first identify a set of plausible hypotheses for $n_1,\ldots,n_K$. This can be performed by inferring clusters based on the spacing between consecutive eigenvalues: if the distance between neighboring eigenvalues is more than a threshold, then we add an entry for a possible cluster separation in the list of all possible positions of cluster separation. From this list, we create all possible K-dimensional vectors of eigenvalue clusters. Obviously, the choice of the threshold is critical to reduce the number of hypotheses to be tested;
2) for each K-dimensional vector with number of antennas $\hat{n}_1,\ldots,\hat{n}_K$, we use Theorem 2 in order to obtain estimates of the $\hat{P}_1,\ldots,\hat{P}_K$ (some being possibly null);
3) based on these estimates, we compare the e.s.d. $F^{\mathbf{B}_N}$ of $\mathbf{B}_N$ to the distribution function $\hat{F}$ defined as the l.s.d. of the matrix model $\hat{\mathbf{Y}}=\mathbf{H}\hat{\mathbf{P}}\mathbf{X}+\mathbf{W}$ with $\hat{\mathbf{P}}$ the diagonal matrix composed of $\hat{n}_1$ entries equal to $\hat{P}_1$, $\hat{n}_2$ entries equal to $\hat{P}_2$ etc. up to $\hat{n}_K$ entries equal to $\hat{P}_K$. The comparison can be performed based on different metrics. In the simulations carried hereafter, we consider as a metric the mean absolute difference between the Stieltjes transform of $F^{\mathbf{B}_N}$ and of $\hat{F}$ on the segment $[-1,-0.1]$.

Note that the above process can bring an interesting feature linked to the cluster separability problem discussed along this article. Indeed, if two subsequent powers P_i and P_{i+1} are close to one another, then the separability condition of Assumptions 1 and 2 is not verified. If one knows n_i and n_{i+1} and blindly uses the estimator of Theorem 2, the result can be catastrophic as the estimator is unreliable. On the contrary, if n_i and n_{i+1} are unknown and one uses the above process, it is very likely that the distinct sources with close power will be assumed to be a single source with power equal to the estimate of $(P_i+P_{i+1})/2$ and embedded with n_i+n_{i+1} antennas. For practical blind detection purposes in cognitive radios, this leads the secondary network to infer a number of transmit entities that is less than the effective number of transmitters. In general, this would not have serious consequences on the decisions made by the secondary network but this might at least reduce the capabilities of the secondary network to optimally overlay the licensed spectrum. Further work is also being carried out to go past the cluster separability assumption; specifically, methods for estimating the number of P_i associated to any cluster j_F are under study.

V. Simulations

In this section, we provide simulation results to assess the performance of Theorem 2 when K, and $n_1,\ldots,n_K$ are known, to compare this performance against alternative estimation methods and finally to evaluate the performance of the *ad-hoc* approach discussed in Section IV-D. In order to underline some precise features of the advantages of our

novel method, we will use two simulation models. The first model, already presented in Figure 2, involves a scenario with clear separation between clusters, while the second model will consider the case of co-located clusters.

The estimator of Theorem 2 will be compared against two methods, which we describe below.

A. Alternative methods

1) Strongly consistent estimator for $M \gg N$ and $N \gg n$: The first method is the classical estimator that assumes that the sample dimension M is much larger than the sensor dimension N, while N is much larger than the source dimension n. In this case, it is easy to see that the e.s.d. of $\mathbf{B}_N$ tends to a mass in σ^2. However, the first n eigenvalues of $\mathbf{B}_N$ are asymptotically greater than σ^2 and it is also clear that the e.s.d. of the projection of $\mathbf{B}_N$ on the eigenspace associated to its largest n eigenvalues tends to K masses in $P_1+\sigma^2, \ldots, P_K+\sigma^2$. This leads to the strongly consistent estimator $\hat{P}_k^{\infty}$ of P_k given by

$$\hat{P}_k^{\infty} = \frac{1}{n_k} \sum_{i \in \mathcal{N}_k} (\lambda_i - \hat{\sigma}^2), \tag{84}$$

with

$$\hat{\sigma}^2 = \frac{1}{N-n} \sum_{i=1}^{N-n} \lambda_i$$

and we recall that $\lambda_1 \leq \ldots \leq \lambda_N$ are the eigenvalues of $\mathbf{B}_N$. The strong consistence is with respect to the rates $n \to \infty$, $N/n \to \infty$ and $M/N \to \infty$. Note that we take an estimator for σ^2 instead of σ^2 itself in order to be coherent with Theorem 2 which does not require any *a priori* information on σ^2. We will refer to this estimator as the *classical* method.

2) Estimator based on strongly consistent moment estimates: The second method is a technique issued from free probability theory, which is based on moments of the l.s.d. of $\mathbf{B}_N$. As such, we will refer to this method as the *moment* method. It consists in computing the first moments of the e.s.d. of $\mathbf{B}_N$, i.e., $\frac{1}{N}\operatorname{tr}\left(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\right)^k$, for $k = 1, \ldots, K$, from which the *deconvolved* moments $\frac{1}{n}(n_1 P_1^k + \ldots + n_K P_K^k)$ of $F^{\mathbf{P}}$ can be evaluated, see e.g., [23]. These estimated moments can be expressed as polynomials of the moments of $F^{\mathbf{B}_N}$, which is convenient from a practical point of view although it leads to serious shortcomings in terms of estimator accuracy. Indeed, small deviations in the low order moments of $F^{\mathbf{B}_N}$ around the corresponding moments of F lead to large deviations in the estimation of the high order moments of $F^{\mathbf{P}}$.

One can then retrieve the vector $(\hat{P}_1^{(\mathrm{mom})}, \ldots, \hat{P}_K^{(\mathrm{mom})})$ whose distribution function has for first K moments the first K estimated moments of $F^{\mathbf{P}}$. This is performed using Newton-Girard polynomial formulas [24], which boils down to finding the roots of a polynomial of order K. The value $\hat{P}_k^{(\mathrm{mom})}$ is the estimate of P_k. Computing $\hat{P}_k^{(\mathrm{mom})}$ requires in particular that $K, n_1, \ldots, n_K$ and σ^2 are known. The main shortcoming of the Newton-Girard inversion is that the polynomial to be solved may have purely imaginary roots. This issue, added to the deviations in the estimated moments, contribute to rather poor estimation accuracies unless the system dimensions are very large. However, as opposed to the classical method and the novel Stieltjes transform approach, the moment method does not require any assumption of cluster separability to be valid.

B. Results

1) Cluster separability limit: We start with a demonstration of the performance of the novel estimator with respect to the satisfaction of the cluster separability assumption. We consider the model presented in Figure 2, i.e., $K = 3$, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$ and $n/N = N/M = 1/10$. The SNR, defined as $\mathrm{SNR} = 1/\sigma^2$, ranges from -15 dB to 20 dB. The entries of $\mathbf{X}$ are QPSK-modulated and those of $\mathbf{H}$ and $\mathbf{W}$ are Gaussian distributed. In Figure 11, we present simulation results in terms of normalized mean square error (NMSE) in the estimates of the individual P_k, both for $n = 60$ and $n = 6$. For future need, we define this system model with $n = 6$ as Scenario (a). The NMSE for power P_k is given by

$$\mathrm{NMSE} = \mathrm{E}\left[\frac{(P_k - \hat{P}_k)^2}{P_k^2}\right], \tag{85}$$

where the expectation is taken over the random realizations of the matrices $\mathbf{H}$, $\mathbf{X}$ and $\mathbf{W}$.

Note how steep the mean square error curves increase below a given SNR value. This intuitively corresponds to the tipping point where the cluster separability assumptions are no longer verified. Especially here, this corresponds to the point where Assumption 2 no longer holds. Now, remembering the results of Figure 6, observe that the horizontal line $c = 10$ crosses the respective curves of validity of Assumption 2 around the SNR values where Figure 11 shows steep curve increase. This indicates that our novel estimator is indeed inappropriate when Assumption 3 is not satisfied. This also validates the accuracy of Assumption 2, which we recall is only a sufficient condition for cluster separability. Note also that, as long as cluster separation is achieved, the performance of the Stieltjes transform algorithm goes quickly down to a constant level (with respect to the SNR) which is a function of the amplitude of the values of n, N and M.

2) Performance comparison: We first compare the classical method against the novel Stieltjes transform approach for Scenario (a). Under the hypotheses of this scenario, the ratios c and c_0 equal 10, leading therefore the classical detector to be almost asymptotically unbiased. We therefore suspect that the NMSE performance for both detectors is alike. This is described in Figure 12, which suggests as predicted that in the high SNR regime (when cluster separability is reached) the classical estimator performs similar to the Stieltjes transform method. However, it appears that a 3 dB gain is achieved by the Stieltjes transform method around the position where cluster separability is no longer satisfied. This translates the fact that, when subsequent clusters tend to merge as σ^2 increases, the Stieltjes transform method manages to track the position of the powers P_k while the classical method keeps assuming each P_k is located at the center of cluster k_F. This observation is very similar to that made in [25], where an improved MUSIC estimator is introduced that pushes further the SNR position

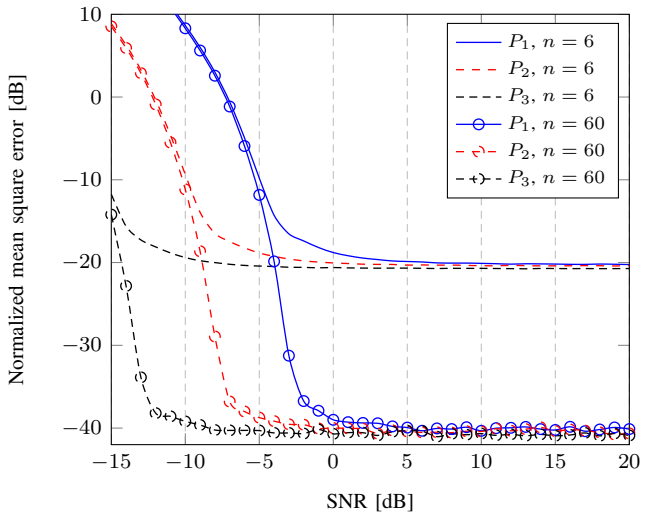

Fig. 11. Normalized mean square error of individual powers $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, $P_1 = 1, P_2 = 3, P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, for $10,000$ simulation runs.

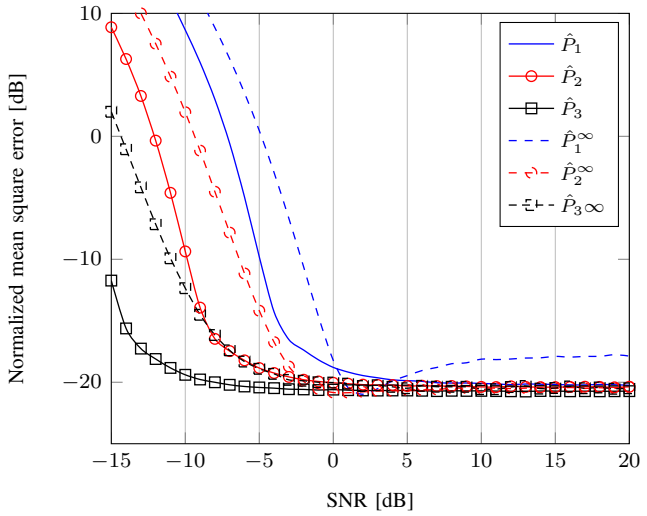

Fig. 12. Normalized mean square error of individual powers $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$, $P_1 = 1, P_2 = 3, P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, $n = 6$. Comparison between classical and Stieltjes transform approach.

where the performance of the classical MUSIC estimator decays significantly.

We now consider another model, for which the classical estimator is largely biased. We now take $K = 3$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1/n = n_2/n = n_3/n = 1/3$ and $n = 12$, $N = 24$ and $M = 128$. The entries of $\mathbf{X}$ are still QPSK-modulated while the entries of $\mathbf{H}$ and $\mathbf{W}$ are still independent standard Gaussian. This model is further referred to as Scenario (b). We first compare the performance of the classical, Stieltjes transform and moment estimators for an SNR of 20 dB. Figure 13 depicts the distribution function of the estimated powers in logarithmic scale. The Stieltjes transform method appears here to be very precise and seemingly unbiased. On the opposite, the classical method, with a slightly smaller variance shows a large bias as was anticipated. As for

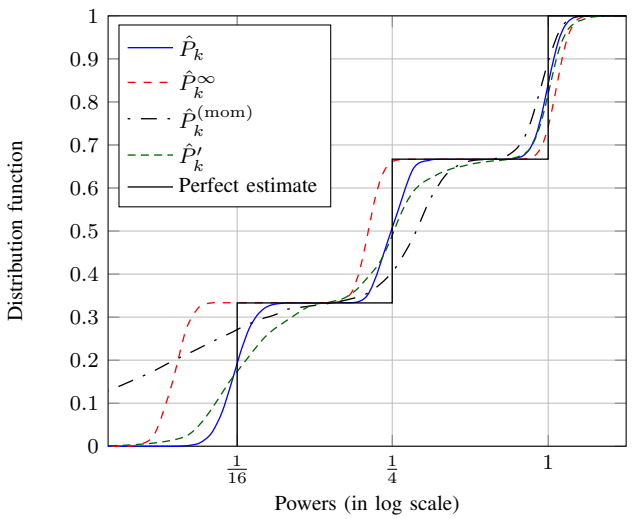

Fig. 13. Distribution function of the estimators $\hat{P}_k^{\infty}$, $\hat{P}_k$, $\hat{P}'_k$ and $\hat{P}_k^{(\mathrm{mom})}$ for $k \in \{1, 2, 3\}$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$ antennas per user, $N = 24$ sensors, $M = 128$ samples and $\mathrm{SNR} = 20$ dB. Optimum estimator shown in dashed lines.

the moment method, it shows rather accurate performance for the stronger estimated power, but proves very inaccurate for smaller powers. This entails from the inherent shortcomings of the moment method. The performance of the estimator $\hat{P}'_k$ will be commented in Section V-B3.

We then focus on the estimate for the larger power P_3 and take now the SNR to range from -15 to 30 dB under the same conditions as previously and for the same estimators. The NMSE for the estimators of P_3 is depicted in Figure 14. The curve marked with squares will be commented in Section V-B3. As already observed in Figure 13, in the high SNR regime, the Stieltjes transform estimator outperforms both alternative methods. We also notice the SNR gain achieved by the Stieltjes transform approach with respect to the classical method in the low SNR regime, as already observed in Figure 12. However, it now turns out that in this low SNR regime, the moment method is gaining ground and outperforms both cluster-based methods. This is due to the cluster separability condition which is not a requirement for the moment approach. This indicates that much can be gained by the Stieltjes transform method in the low SNR regime if a more precise treatment of overlapping clusters is taken into account.

3) Joint estimation of K, n_k, P_k: So far, we have assumed that the number of users K and the number of antennas per user n_k were perfectly known. As discussed in Section IV-D, this may not be a strong assumption if it is known by advance how many antennas are systematically used by every source or if another mechanism, such as in [9], can provide this information. Nonetheless, these are in general strong assumptions to take. Based on the *ad-hoc* method described in Section IV-D, we therefore provide the performance of our novel Stieltjes transform method in the high SNR regime when only n is known; this assumption is less stringent as in the medium to high SNR regime, one can easily decide which eigenvalues of $\mathbf{B}_N$ belong to the cluster associated to σ^2

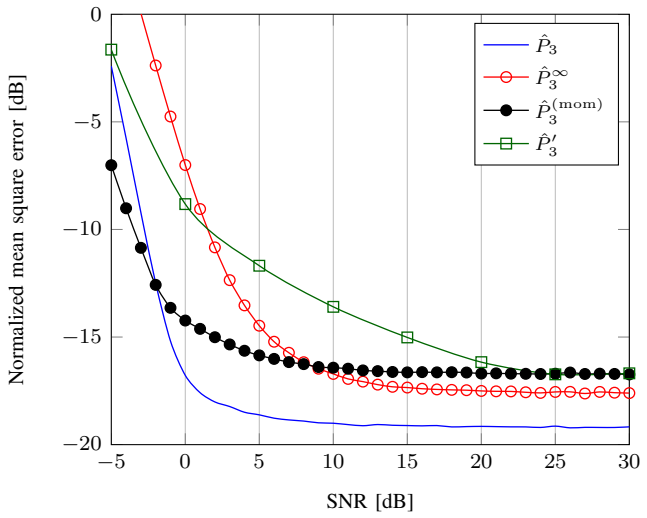

Fig. 14. Normalized mean square error of largest estimated power P_3, $P_1 = 1/16, P_2 = 1/4, P_3 = 1$, $n_1 = n_2 = n_3 = 4$,$N = 24$, $M = 128$. Comparison between classical, moment and Stieltjes transform approaches.

and which eigenvalues do not. We denote $\hat{P}_k'$ the estimator of P_k when K and $n_1, \ldots, n_K$ are unknown. We assume for this estimator that all possible combinations of 1 to 3 clusters can be generated from the $n = 6$ observed eigenvalues in Scenario (a) and that all possible combinations of 1 to 3 clusters with even cluster size can be generated from the $n = 12$ eigenvalues of $\mathbf{B}_N$ in Scenario (b). For Scenario (a), the NMSE performance of the estimators $\hat{P}_k$ and $\hat{P}_k'$ is proposed in Figure 15 for the SNR ranging from 5 dB to 30 dB. For Scenario (b), the distribution function of the inferred $\hat{P}_k'$ is depicted in Figure 13, while the NMSE performance for the inference of P_3 is proposed in Figure 14; these are both compared against the classical, moment and Stieltjes transform estimator. We also indicate in Table V-B3 the percentage of correct estimation of the triplet (n_1, n_2, n_3) for both Scenario (a) and (b). In Scenario (a), this amounts to 12 such triplets that satisfy $n_k \geq 0$, $n_1 + n_2 + n_3 = 6$, while in Scenario (b), this corresponds to 16 triplets that satisfy $n_k \in 2\mathbb{N}$, $n_1 + n_2 + n_3 = 12$. Observe that the noise variance, assumed to be known *a priori* in this case, plays an important role with respect to the statistical inference of the n_k. In Scenario (a), for an SNR greater than 15 dB, the correct hypothesis for the n_k is almost always taken and the performance of the estimator is similar to that of the optimal estimator. In Scenario (b), the detection of the exact cluster separation is less accurate and the performance for the inference of P_3 saturates at high SNR to -16 dB of NMSE, against -19 dB when the exact cluster separation is known. It therefore seems that in the high SNR regime the performance of the Stieltjes transform detector is loosely affected by the absence of knowledge about the cluster separation. This statement is also confirmed by the distribution function of $\hat{P}_k'$ in Figure 13, which still outperforms the classical and moment methods. We underline again here that this is merely the result of an *ad-hoc* approach; this performance could be greatly improved if e.g., more is known about the second order statistics of $F^{\mathbf{B}_N}$.

SNR	RCI (a)	RCI (b)
5 dB	0.8473	0.1339
10 dB	0.9026	0.4798
15 dB	0.9872	0.4819
20 dB	0.9910	0.5122
25 dB	0.9892	0.5455
30 dB	0.9923	0.5490

TABLE I
RATE OF CORRECT INFERENCE (RCI) OF THE TRIPLET (n_1, n_2, n_3) FOR SCENARIOS (A) AND (B).

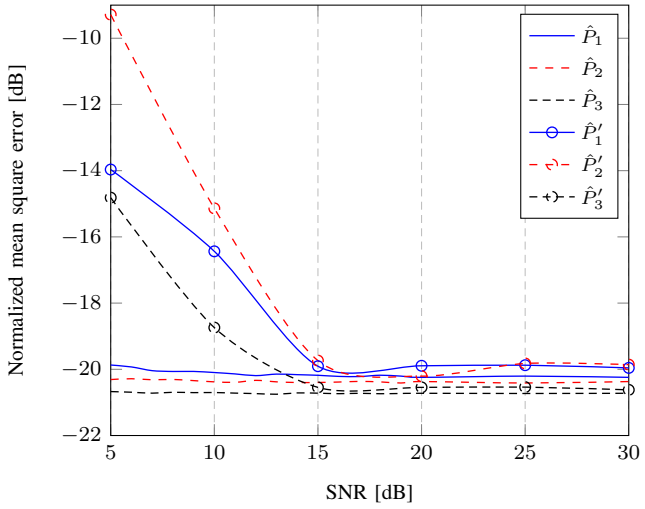

Fig. 15. Normalized mean square error of individual powers $\hat{P}_1$, $\hat{P}_2$, $\hat{P}_3$ and $\hat{P}_1'$, $\hat{P}_2'$, $\hat{P}_3'$, $P_1 = 1, P_2 = 3, P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, $n = 6$, $10,000$ simulation runs.

VI. CONCLUSION

In this paper, a blind multi-source power estimator was derived. Under the assumptions that the ratio between the number of sensors and the number of signals sources is not too small and the source transmit powers are sufficiently distinct from one another, we derived a method to infer the individual source powers if the number of sources is known, which was shown to outperform alternative estimation techniques in the medium to high SNR regime. We then briefly discussed the joint estimation of the number of transmit sources, the number of antennas of each source and the transmit powers, which appeared in simulation to perform well in the high SNR regime. The novel method is moreover computationally efficient and is particularly robust to small system dimensions. As such, it is particularly suited to the blind detection of primary mobile user in future cognitive radio networks.

APPENDIX A
RESIDUE CALCULUS

The integrand of $\hat{P}_k$ in (63) can be expanded as

$$
\begin{aligned}
&-\frac{N}{n}\frac{1}{z\hat{m}_{\underline{F}}(z)}-\frac{N}{n}\frac{\hat{m}'_{\underline{F}}(z)}{\hat{m}_{\underline{F}}(z)^2}-\frac{N}{n}\frac{\hat{m}'_F(z)}{\hat{m}_{\underline{F}}(z)\hat{m}_F(z)}\\
&-\frac{N-n}{n}\frac{1}{z^2\hat{m}_{\underline{F}}(z)\hat{m}_F(z)}-\frac{N-n}{n}\frac{\hat{m}'_{\underline{F}}(z)}{z\hat{m}_{\underline{F}}(z)^2\hat{m}_F(z)}\\
&-\frac{N-n}{n}\frac{\hat{m}'_F(z)}{z\hat{m}_F(z)^2\hat{m}_{\underline{F}}(z)}-\frac{N\sigma^2}{n}\frac{1}{z}-\frac{N\sigma^2}{n}\frac{\hat{m}'_{\underline{F}}(z)}{\hat{m}_{\underline{F}}(z)}\\
&-\frac{N\sigma^2}{n}\frac{\hat{m}'_F(z)}{\hat{m}_F(z)}
\end{aligned}
\tag{86}
$$

First assume the case $M \neq N$. Numbering the nine terms in order, we have that (1) has poles in $z \in \{\eta_1, \ldots, \eta_N\}$, where $\hat{m}_{\underline{F}}(z) = 0$. Applying l'Hospital rule, all poles have order 1 and the corresponding residues are

$$
\lim_{z\to\eta_i} -\frac{N}{n}\frac{z-\eta_i}{z\hat{m}_{\underline{F}}(z)} = -\frac{N}{n}\frac{1}{\eta_i\hat{m}'_{\underline{F}}(\eta_i)} \tag{87}
$$

As for (2), it is the derivative of $-1/\hat{m}_{\underline{F}}(z)$, which is well-behaved inside $\mathcal{C}_{F,k}$, so it does not have poles. The term (3) has poles of order 1 in $z \in \{\eta_1, \ldots, \eta_N\}$ as well and we have the residue

$$
\lim_{z\to\eta_i} -\frac{N}{n}\frac{(z-\eta_i)\hat{m}'_F(z)}{\hat{m}_{\underline{F}}(z)\hat{m}_F(z)} = -\frac{N}{n}\frac{N}{M-N}\frac{\hat{m}'_F(\eta_k)}{\hat{m}'_{\underline{F}}(\eta_k)}\eta_k \tag{88}
$$

the last equality being obtained from the fact that

$$
\hat{m}_F(z) = \frac{M}{N}\hat{m}_{\underline{F}}(z) + \frac{M-N}{N}\frac{1}{z} \tag{89}
$$

and $\hat{m}_F(\eta_i) = 0$. It also has poles in $z \in \{\mu_1, \ldots, \mu_N\}$, where $\hat{m}_F(z) = 0$. These are order 1 poles and we have

$$
\lim_{z\to\mu_i} -\frac{N}{n}\frac{(z-\mu_i)\hat{m}'_F(z)}{\hat{m}_{\underline{F}}(z)\hat{m}_F(z)} = \frac{N}{n}\frac{M}{M-N}\mu_k \tag{90}
$$

Term (4) is shown in a similar way to have residues $-\frac{N-n}{n}\frac{N}{M-N}\frac{1}{\eta_i}\hat{m}'_{\underline{F}}(\eta_i)$ and $\frac{N-n}{n}\frac{M}{M-N}\frac{1}{\mu_i}\hat{m}'_F(\mu_i)$, $i \in \{1, \ldots, N\}$. Term (5) has residues $\frac{N-n}{n}\frac{NM}{(M-N)^2}\eta_i$ and $-\frac{N-n}{n}\frac{M^2}{(M-N)^2}\frac{\mu_i\hat{m}'_F(\mu_i)}{\hat{m}'_F(\mu_k)}$, $i \in \{1, \ldots, N\}$. Term (6) has residues $-\frac{N-n}{n}\frac{N^2}{(M-N)^2}\frac{\eta_i\hat{m}'_F(\eta_i)}{\hat{m}'_{\underline{F}}(\eta_i)}$ and $\frac{N-n}{n}\frac{MN}{(M-N)^2}\mu_i$, $i \in \{1, \ldots, N\}$. Term (7) has a pole in $z = 0$ but we already know that 0 is not inside $\mathcal{C}_{F,k}$, so this is already discarded. Term (8) has poles in $z \in \{\lambda_1, \ldots, \lambda_N\}$ of residue $\frac{N\sigma^2}{n}$ and poles in $z \in \{\eta_1, \ldots, \eta_N\}$ of residue $-\frac{N\sigma^2}{n}$. Similarly term (9) has poles in $z \in \{\lambda_1, \ldots, \lambda_N\}$ of residue $\frac{N\sigma^2}{n}$ and poles in $z \in \{\mu_1, \ldots, \mu_N\}$ of residue $-\frac{N\sigma^2}{n}$.

Summing together the 9 terms and remarking that

$$
\frac{N}{M-N}\hat{m}'_F(z) = \frac{M}{M-N}\hat{m}'_{\underline{F}}(z) - \frac{1}{z^2} \tag{91}
$$

we obtain exactly (64).

Assume now $M = N$, in which case $\hat{m}_{\underline{F}}(z) = \hat{m}_F(z)$. It can be readily seen that the terms (4) to (6) are the derivative of $-\frac{N-n}{n}\frac{1}{z^2\hat{m}_F(z)}$, so that they have residue 0. The only remaining term here is (1), whose residues are the $-\frac{N}{n}\frac{1}{\eta_i\hat{m}'_F(\eta_i)}$.

APPENDIX B
PROOF OF EXACT SEPARATION

Theorem 3 is a generalization from the assumption of identical distribution of the x_{ij}'s, the proof of which is contained in the two papers [21] and [17], which, with some modifications, appear as Chapter 6 in [16]. The proof uses previous articles that need to be updated as well. We shall therefore go through the necessary steps that need to be modified, taking reference to all papers successively.

We shall assume for simplicity in the following that the matrices $\mathbf{T}_n$ are deterministic, converges in distribution to H and that $\|\mathbf{T}_n\|$ is uniformly bounded. The generalization to random $\mathbf{T}_n$ follows from Tonelli's theorem [20]. Indeed, let $\mathcal{X}$ be the probability space that engenders the $\mathbf{X}_n$ and $\mathcal{T}$ the probability space that engenders the $\mathbf{T}_n$. Let A be any of the events in Conclusion 1), 2), or 3), claimed to occur with probability one. Assume that Theorem 3 holds for deterministic $\mathbf{T}_n$ satisfying Assumptions 5), 6), and 7). Let $t \in \mathcal{T}$ be an element of the intersection of these events. Then $I_A(t, x) = 1$ for all x contained in a subset of $\mathcal{X}$ having probability one. Therefore, by Tonelli's theorem, denoting $\mathcal{T} \times \mathcal{X}$ the product space of $\mathcal{T}$ and $\mathcal{X}$, we have that

$$
\begin{aligned}
&\int_{\mathcal{T}\times\mathcal{X}} I_A(t,x)dP_{\mathcal{T}\times\mathcal{X}}(t,x)\\
&\quad= \int_{\mathcal{T}}\left[\int_{\mathcal{X}} I_A(t,x)dP_{\mathcal{X}}(x)\right]dP_{\mathcal{T}}(t) = 1,
\end{aligned}
\tag{92}
$$

and Theorem 3 therefore holds true if it holds true for $\mathbf{T}_n$ deterministic.

A. Extension of [26]

The first step is to extend the work in [26] on the largest eigenvalue of $\mathbf{S}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^{\mathsf{H}}$, where $\mathbf{X}_n = (x_{ij})$ is $p \times n$, $p = p(n)$, and $p/n \to y > 0$ as $n \to \infty$. Checking the assumptions in [26], we change the six conditions of Page 518 to

(1) $x_{ij}, i = 1, 2, \ldots, p;\ j = 1, 2, \ldots, n$ are independent for each n,
(2) $|x_{ij}| < \eta_n\sqrt{n}$, where $\eta_n \downarrow 0$,
(3) $\mathrm{E}x_{ij} = 0$,
(4) $\mathrm{E}|x_{ij}^2| \leq 1$,
(5) $\mathrm{E}|x_{ij}|^l < (\eta_n\sqrt{n})^{l-1}$, for $l \geq 2$,
(6) $\mathrm{E}|x_{ij}^l| \leq c(\eta_n\sqrt{n})^{l-3}$, for $l \geq 3$.

By the same argument given there (no difference for complex random variables), the inequality

$$
\mathrm{E}\,\mathrm{tr}(\mathbf{S}_n)^k \leq \eta^k \tag{93}
$$

holds for any $\eta > b \equiv (1 + \sqrt{y})^2$ provided k is chosen such that

(a) $k/\log n \to \infty$,
(b) $\eta_n^{\frac{1}{6}}k/\log n \to 0$.

This implies that $\mathbf{P}(\lambda_{\max}(\mathbf{S}_n) > b + \varepsilon) = o(n^{-t})$ for any given $\varepsilon > 0$ and $t > 0$.

Remark 3: Notice that if Condition (4) is replaced by $\mathrm{E}|x_{ij}^2| \leq \iota$, where ι is a fixed positive constant, then we have

$$
\mathbf{P}(\lambda_{\max}(\mathbf{S}_n) > \iota(b + \varepsilon)) = o(n^{-t}). \tag{94}
$$

We only need to consider the matrix $\iota^{-1}\mathbf{S}_n$ and replace x_{ij} by $\iota^{-1/2}x_{ij}$ to verify the six conditions.

B. First step truncation and renormalization

We consider $\mathbf{S}_n\mathbf{T}_n$, where the assumptions of Theorem 3 are met, except the $\mathbf{T}_n$ are assumed nonrandom. Here, to be consistent with Chapter 6 of [16], we replace c_n with y_n and c with y.

Notice, from the identity

$$\mathrm{E}Y^4 = \int_0^\infty P(Y > x^{1/4})dx, \tag{95}$$

valid for any nonnegative random variable Y, that (ii) implies the fourth moments of the x_{ij} exist.

We will need the following identity later on. For nonnegative Y having finite fourth moment, since $\mathrm{E}Y^4 I(Y > y) \geq y^4 P(Y > y)$, we have $y^4 P(Y > y) \to 0$ as $y \to \infty$. Thus, using integration by parts, we have for any $a > 0$

$$\begin{aligned}
&\mathrm{E}Y^4 I(Y > a) \\
&= a^4 P(Y > a) + \lim_{y\to\infty}(-y^4 P(Y > y) + \int_a^y 4x^3 P(Y > x)dx) \qquad (96)\\
&= a^4 P(Y > a) + \int_a^\infty 4x^3 P(Y > x)dx. \qquad (97)
\end{aligned}$$

We choose $\eta_n \downarrow 0$ such that $\eta_n\sqrt{n} \uparrow \infty$, $\liminf_n \eta_n^2\sqrt{n} > 0$, and

$$\sum_{k=1}^\infty 2^{2k} P(|X| > \tilde{\eta}_k 2^{k/2}) < \infty, \tag{98}$$

where $\tilde{\eta}_k = \eta_{2^k}$.

1. Truncation. Define $y_{ij} = x_{ij} I(|x_{ij}| \leq \eta\sqrt{n})$, $\mathbf{Y}_n = (y_{ij})_{p\times n}$ and $\widehat{\mathbf{S}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}$. We have

$$\begin{aligned}
&P(\mathbf{X}_n \neq \mathbf{Y}_n, \text{ i.o.}) \\
&\leq \lim_{m\to\infty}\sum_{k=m}^\infty P\left(\bigcup_{n=2^k+1}^{2^{k+1}}\bigcup_{i\leq p, j\leq n}\{|x_{ij}| > \eta_n\sqrt{n}\}\right) \qquad (99)\\
&\leq \lim_{m\to\infty}\sum_{k=m}^\infty P\left(\bigcup_{n=2^k+1}^{2^{k+1}}\bigcup_{i\leq 2yn, j\leq n}\{|x_{ij}| > \tilde{\eta}_k 2^{k/2}\}\right) \qquad (100)\\
&= \lim_{m\to\infty}\sum_{k=m}^\infty P\left(\bigcup_{i\leq y2^{k+2}, j\leq 2^{k+1}}\{|x_{ij}| > \tilde{\eta}_k 2^{k/2}\}\right) \qquad (101)\\
&= 8yK\lim_{m\to\infty}\sum_{k=m}^\infty 2^{2k} P\left(|X| > \tilde{\eta}_k 2^{k/2}\right) = 0. \qquad (102)
\end{aligned}$$

2. Centralization. Define $z_{ij} = y_{ij} - \mathrm{E}y_{ij}$ and $\mathbf{Z}_n = (z_{ij})_{p\times n}$ and $\widetilde{\mathbf{S}}_n = \frac{1}{n}\mathbf{Z}_n\mathbf{Z}_n^{\mathsf{H}}$. Then by Theorem A. 46 of [16] and the above identity, we have

$$\begin{aligned}
&\max_{k\leq p}|\lambda_k^{\frac{1}{2}}(\widehat{\mathbf{S}}_n\mathbf{T}_n) - \lambda_k^{\frac{1}{2}}(\widetilde{\mathbf{S}}_n\mathbf{T}_n)| \\
&\leq \|\mathbf{T}_n^{\frac{1}{2}}\|\|n^{-\frac{1}{2}}\mathrm{E}(\mathbf{Y}_n)\| \qquad (103)\\
&\leq \left(\frac{1}{n}\sum_{ij}|\mathrm{E}y_{ij}|^2\right)^{\frac{1}{2}} \qquad (104)\\
&= \left(\frac{1}{n}\sum_{ij}|\mathrm{E}x_{ij}I(|x_{ij}| > \eta_n\sqrt{n})|^2\right)^{\frac{1}{2}} \qquad (105)\\
&\leq \left(\frac{1}{\eta_n^4 n^3}\sum_{ij}\mathrm{E}|x_{ij}^2|\mathrm{E}|x_{ij}^4|I(|x_{ij}| > \eta_n\sqrt{n})\right)^{\frac{1}{2}} \qquad (106)\\
&\leq \left(\frac{Knp}{\eta_n^4 n^3}\mathrm{E}|X^4|I(|X| > \eta_n\sqrt{n})\right)^{\frac{1}{2}} \to 0. \qquad (107)
\end{aligned}$$

3. Rescaling. Define $w_{ij} = z_{ij}/\sigma_{ij}$, $\mathbf{W}_n = (w_{ij})_{p\times n}$ and $\breve{\mathbf{S}}_n = \frac{1}{n}\mathbf{W}\mathbf{W}_n^{\mathsf{H}}$. Then, by Theorem A. 46 of [16],

$$\begin{aligned}
&\max_{k\leq p}|\lambda_k^{\frac{1}{2}}(\breve{\mathbf{S}}_n\mathbf{T}_n) - \lambda_k^{\frac{1}{2}}(\widetilde{\mathbf{S}}_n\mathbf{T}_n)| \\
&\leq \|\mathbf{T}_n^{\frac{1}{2}}\|\|n^{-\frac{1}{2}}(\mathbf{Z}_n - \mathbf{W}_n)\| \qquad (108)\\
&\leq \left\|\frac{1}{\sqrt{n}}\left[z_{ij}(1-\lambda_{ij}^{-1})\right]\right\| \to 0, \text{ a.s.} \qquad (109)
\end{aligned}$$

because of (94) and the fact that

$$\begin{aligned}
&\max_{i,j}|1-\lambda_{ij}^2| \qquad (110)\\
&\leq \max_{ij}[\mathrm{E}|x_{ij}^2|I(|x_{ij}| > \eta\sqrt{n}) + (\mathrm{E}|x_{ij}|I(|x_{ij}| > \eta\sqrt{n}))^2] \qquad (111)\\
&\leq 2\psi^{-1}(\eta\sqrt{n})\max_{ij}\mathrm{E}|x_{ij}^2|\psi(|x_{ij}|) \to 0 \qquad (112)
\end{aligned}$$

which implies that

$$\max_{ij}\mathrm{E}|w_{ij} - z_{ij}|^2 = \max_{ij}\frac{(1-\lambda_{ij}^2)^2}{(1+\lambda_{ij})^2} \to 0. \tag{113}$$

C. Second truncation and normalization

We may assume now that the x_{ij} satisfy the six conditions of Section B-A with Condition (4) strengthened to $\mathrm{E}|x_{ij}^2| = 1$ for all i, j.

Define $y_{ij} = x_{ij}I(|x_{ij}| \leq C) - \mathrm{E}x_{ij}I(|x_{ij}| \leq C)$ for some large constant C and define $\mathbf{Y}_n = (y_{ij})_{p\times n}$, $\widehat{\mathbf{S}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}$.

Then by Theorem A. 46 of [16], we have

$$\begin{aligned}
&\max_{k\leq p}|\lambda_k^{\frac{1}{2}}(\widehat{\mathbf{S}}_n\mathbf{T}_n) - \lambda_k^{\frac{1}{2}}(\widetilde{\mathbf{S}}_n\mathbf{T}_n)| \\
&\leq \|\mathbf{T}_n^{\frac{1}{2}}\|\left\|n^{-1/2}(\mathbf{X}_n - \mathbf{Y}_n)\right\| \qquad (114)\\
&\leq \left\|n^{-1/2}(\mathbf{X}_n - \mathbf{Y}_n)\right\|. \qquad (115)
\end{aligned}$$

Since $\mathrm{E}|x_{ij} - y_{ij}|^2 \leq \mathrm{E}|x_{ij}^2|I(|x_{ij}| > C) \leq M/\psi(C)$, this can be made arbitrarily small by making C sufficiently large. We can then apply (94).

The rescaling is the same as given in last Section. Now we have

$$\max_{ij} |1-\sigma_{ij}^2| \leq 2\psi^{-1}(C) \max_{ij} \mathrm{E}|x_{ij}^2|\psi(x_{ij}), \tag{116}$$

which can be made arbitrarily small by making C sufficiently large.

D. Extension of [27] and Chapter 6 of [16]

The result in [27] on the smallest eigenvalue, $\lambda_{\min}(\mathbf{S}_n)$, of $\mathbf{S}_n$ can be extended with only Assumptions 1), 2), 3) of Theorem 3. Indeed, using the two step truncations, we may assume the x_{ij} are bounded, with mean 0 and variance 1. Following the same steps as in [27], one may prove that when $y<1$

$$\lambda_{\min}(\mathbf{S}_n) \to (1-\sqrt{y})^2, \tag{117}$$

almost surely.

We proceed now to the necessary changes in Chapter 6 in [16]. We may now assume the same conditions as in Section 6.2.1 of [16] on the x_{ij} (except they need not be identically distributed), and the bounds appearing there. The changes are needed wherever identical distribution was exploited.

We begin with Page 139, below (6.2.34). We change the definition of b_n to

$$b_n = \frac{1}{1+n^{-1}\mathrm{E}\,\mathrm{tr}(\mathbf{T}_n\mathbf{D}^{-1})}, \tag{118}$$

and introduce the quantities

$$b_{nj} = \frac{1}{1+n^{-1}\mathrm{E}\,\mathrm{tr}(\mathbf{T}_n\mathbf{D}_j^{-1})}. \tag{119}$$

The argument below (6.2.35) is specific for $j=1$, but easily extends for any j. Following the argument below (6.2.36), we can no longer assume $\mathrm{E}\beta_1 = -z\mathrm{E}\underline{s}_n$, nor is bounded, but we have

$$\sup_{u\in[a,b]} \left| \frac{1}{n}\sum_k \mathrm{E}\beta_k \right| \leq K. \tag{120}$$

We further have

$$b_{nk} = \beta_k + \beta_k b_{nk}\gamma_k. \tag{121}$$

Then, using (6.2.36)

$$\frac{1}{n}\left|\sum_k (b_{nk}-\mathrm{E}\beta_k)\right| \leq Kn^{-1}\sum_k v_n^{-2}(\mathrm{E}|\gamma_k|^2)^{\frac{1}{2}} \tag{122}$$

$$\leq v_n^{-3}n^{-\frac{1}{2}}. \tag{123}$$

Since $b_{nj}-b_n = b_n b_{nj}\mathrm{E}(\frac{1}{n}\mathrm{tr}\,\mathbf{T}(\mathbf{D}^{-1}-\mathbf{D}_j^{-1}))$, we have, using Lemma 6.9 of [16]

$$\left| b_n - \frac{1}{n}\sum_k b_{nk}\right| = \frac{1}{n}\left|\sum_k (b_n-b_{nk})\right| \leq |z|^2\frac{1}{nv^3}, \tag{124}$$

and

$$|b_{nk}-b_n| \leq K\frac{1}{nv^3}. \tag{125}$$

Thus we have

$$\max_j \sup_{u\in[a,b]} |b_{nj}| \leq K. \tag{126}$$

For the rest of Section 6.2.3, b_n is mentioned twice. We need to replace it with b_{nj} and the arguments go through without any further changes.

For Section 6.2.4, (6.2.42) needs to be replaced by

$$y_n \int \frac{dH_n(t)}{1+t\mathrm{E}\underline{s}_n} + zy_n\mathrm{E}(s_n(z)) \tag{127}$$

$$= \frac{1}{n}\sum_{k=1}^n \mathrm{E}\beta_k \left[\mathbf{r}_k^{\mathsf{H}}\mathbf{D}_k^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{r}_k - \frac{1}{n}\mathrm{E}\,\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}^{-1}\right]. \tag{128}$$

For the rest of the section, replace subscript 1 with subscript k, subscript 2 with subscript j, replace b_{1n} with

$$b_{kj} = \frac{1}{1+n^{-1}\mathrm{E}\,\mathrm{tr}(\mathbf{T}_n\mathbf{D}_{kj})^{-1}},\ k\neq j, \tag{129}$$

and all appearances of subscripts kj assume $k\neq j$. F_{nkj} has the obvious definition. Replace the summations for j ranging from 2 to n with $j\neq k$. All the bounds derived for $k=1$ are true for all k. So we conclude the left side of (6.2.42) is bounded by kn^{-1}.

The rest of Chapter 6 follows without any changes.

REFERENCES

[1] J. M. III and G. Q. M. Jr, "Cognitive radio: making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, 1999.

[2] H. Claussen, L. T. Ho, and L. G. Samuel, "An overview of the femtocell concept," *Bell Labs Technical Journal*, vol. 13, no. 1, pp. 221–245, May 2008.

[3] D. Calin, H. Claussen, and H. Uzunalioglu, "On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 26–32, Jan. 2010.

[4] V. Chandrasekhar, M. Kountouris, and J. G. Andrews, "Coverage in Multi-Antenna Two-Tier Networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5314–5327, 2009.

[5] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[6] V. I. Kostylev, "Energy detection of a signal with Random Amplitude," in *Proc. IEEE International Conference on Communications (ICC'02)*, New York, NY, USA, 2002, pp. 1606–1610.

[7] R. Couillet and M. Debbah, "A Bayesian framework for collaborative multi-source signal detection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5186–5195, Oct. 2010.

[8] P. Bianchi, J. Najim, M. Maida, and M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," *IEEE Trans. Inf. Theory*, 2010, to appear.

[9] P. Chung, J. Böhme, C. Mecklenbraüker, and A. Hero, "Detection of the Number of Signals Using the Benjamini-Hochberg Procedure," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2497–2508, 2007.

[10] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[11] J. W. Silverstein and P. L. Combettes, "Signal detection via spectral theory of large dimensional random matrices," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2100–2105, 1992.

[12] N. E. Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Annals of Statistics*, vol. 36, no. 6, pp. 2757–2790, Dec. 2008.

[13] N. R. Rao, J. A. Mingo, R. Speicher, and A. Edelman, "Statistical eigen-inference from large Wishart matrices," *Annals of Statistics*, vol. 36, no. 6, pp. 2850–2885, Dec. 2008.

[14] R. Couillet and M. Debbah, "Free deconvolution for OFDM multicell SNR detection," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, France, 2008.

[15] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.
[16] Z. Bai and J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," *Springer Series in Statistics*, 2009.
[17] Z. D. Bai and J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.
[18] W. Rudin, *Real and complex analysis*, 3rd ed. McGraw-Hill Series in Higher Mathematics, May 1986.
[19] J. W. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.
[20] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.
[21] Z. D. Bai and J. W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.
[22] D. Gregoratti and X. Mestre, "Random DS/CDMA for the amplify and forward relay channel," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 1017–1027, 2009.
[23] O. . Ryan and M. Debbah, "Free deconvolution for signal processing applications," in *Proc. IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, Jun. 2007, pp. 1846–1850.
[24] R. Séroul, *Programming for Mathematicians*. New York, NY, USA: Springer Universitext, Feb. 2000.
[25] X. Mestre and M. Lagunas, "Modified Subspace Algorithms for DoA Estimation With Large Arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–614, Feb. 2008.
[26] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability Theory and Related Fields*, vol. 78, no. 4, pp. 509–521, 1988.
[27] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *The Annals of Probability*, vol. 21, no. 3, pp. 1275–1294, 1993.

Romain Couillet was born in Abbeville, France. He received his Msc. in Mobile Communications at the Eurecom Institute, France in 2007. He received his Msc. in Communication Systems in Telecom ParisTech, France in 2007. In September 2007, he joined ST-Ericsson (formerly NXP Semiconductors, founded by Philips). At ST-Ericsson, he works as an Algorithm Development Engineer on the Long Term Evolution Advanced (LTE-A) project. In parallel to his position at ST-Ericsson, he is currently a PhD student at Suplec, France. His research topics include mobile communications, multi-users multi-antenna detection, cognitive radio cognitive, Bayesian probability and random matrix theory. He is the recipient of the ValueTools best student paper award, 2008.

Jack W. silverstein received the B.A. degree in mathematics from Hofstra University, Hempstead, NY, in 1971 and the M.S. and Ph.D. degrees in applied mathematics from Brown University, Providence, RI, in 1973 and 1975, respectively. After postdoctoring and teaching at Brown, he began in 1978 a tenure track position in the Department of Mathematics at North Carolina State University, Raleigh, where he has been Professor since 1994. His research interests are in probability theory with emphasis on the spectral behavior of large-dimensional random matrices. Prof. Silverstein was elected Fellow of the Institute of Mathematical Statistics in 2007.

Mérouane Debbah was born in Madrid, Spain. He entered the Ecole Normale Suprieure de Cachan (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively in 1999 and 2002. From 1999 to 2002, he worked for Motorola Labs on Wireless Local Area Networks and prospective fourth generation systems. From 2002 until 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (FTW) (Vienna, Austria). From 2003 until 2007, he joined the Mobile Communications de-partment of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. He is presently a Professor at Supelec (Gif-sur-Yvette, France), holder of the Alcatel-Lucent Chair on Flexible Radio. His research interests are in information theory, signal processing and wireless communications. Mrouane Debbah is the recipient of the "Mario Boella" prize award in 2005, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the2010 Newcom++ best paper award as well as the Valuetools 2007,Valuetools 2008 and CrownCom2009 best student paper awards. He is a WWRF fellow.

Zhidong Bai graduated from University of Science and Technology of China, majoring statistics and probability. His research interests include limiting theorems of statistics and spectral analysis of large dimensional random matrices, rounded data analysis etc. He is currently a professor of the School of Mathematics and Statistics at Northeast Normal University, China, and Department of Statistics and Applied Probability at National University of Singapore. He is a Fellow of the Third World Academy of Sciences and a Fellow of the Institute of Mathematical Statistics.